THE UNIVERSITY
of EDINBURGH

Disfluency, prediction and attention in language comprehension.

Samuel Miller

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2015

DECLARATION

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

SAMUEL J. MILLER

Abstract

Spoken language comprehension is impacted by the presence of disfluencies. It follows that there have been attempts to understand the underlying mechanisms that are responsible for these disfluency effects. Different accounts of disfluency processing have been proposed to explain these effects; the current thesis was directed towards exploring two standpoints of disfluency processing: the predictional and attentional accounts.

Disfluency has been shown to modulate predictive processing, with a clear effect in the literature being that upon encountering disfluency listeners show a bias for unknown or discourse new referents (Arnold, Kam, & Tanenhaus, 2007; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Heller, Arnold, Klein, & Tanenhaus, 2014). The predictional account interprets this finding in terms of expectancy: according to this view, listeners expect speakers to produce harder-to-access words in situations where their linguistic performance is consistent with planning problems.

Listeners are also more likely to remember words that follow a disfluency (Corley et al., 2007). The presence of disfluency has been shown to affect the attentional state of the listener, as indexed by attenuation of event related potentials to acoustically manipulated words post disfluency (Collard et al., 2008). These effects form the basis of the competing attentional account, which suggests that that upon encountering disfluency, listeners stop predicting about upcoming content and instead, employ heightened attentional resources to help them resolve the situation.

In the first experiment, we aimed to distinguish between the predictional and attentional accounts by employing a visual world paradigm to investigate directly the underlying mechanism during comprehension. Participants were expected to show different fixation behaviour depending on which account was true. The main experiment provided some unexpected results, as the fixation behaviour seen would

not have been predicted by either account. These results were further investigated in a number of post-hoc tests, testing participants sensitivity to the disfluency used in the main paradigm. The results observed were again inconclusive. Taken together these findings suggested that the mechanisms afforded by each account for disfluency processing may work in unison, with reliance on either attentional or predictive processing, or a mix of both, dictated by the demands of the task.

In the remaining experiments (2-6) we focused on the attentional account of disfluency processing; we asked how disfluencies impact listener attention at a phonemic level. Pitt and Szostak (2012) demonstrated that the effect of phoneme manipulation is reduced when participants' attention is explicitly directed to the ambiguous phoneme, with participants less likely to categorise an ambiguous item as a "word" under such conditions than otherwise. We applied this paradigm at the sentence level to investigate whether disfluencies induce heightened attentional focus at a phonemic level. Specifically, we compared the impact of a phoneme manipulation on lexicality judgements with; (i) attentional focus, and; (ii) disfluency presence. The initial experiments' findings failed to replicate the attentional manipulation seen in the Pitt and Szostak study (2012) but results from the later studies suggested there is evidence that disfluency does drive listener attention but actually makes listeners more accommodating of the phoneme manipulation heard. These results are discussed in relation to the accounts of disfluency processing being tested.

Acknowledgements

I would like to acknowledge the ESRC for their studentship that funded me throughout my PhD, without which the research needed for the current thesis would not have been possible. The University of Edinburgh, School of Philosophy, Psychology and Language Sciences are also worthy of note for their funding of crucial parts of the research process and accompanying conference fees for studies contained in the current thesis.

I would like to express my continued gratitude for the support afforded to me by a number of individuals. First, a great deal of thanks towards my principal supervisor, Dr Martin Corley for his invaluable support and guidance during the undertaking of the current research. Next, a similar deal of thanks goes to my second supervisor Professor Martin Pickering for his assistance in all matters thesis related. Their joint enthusiasm, knowledge and patience when discussing, advising and correcting drafts at all stages for the current thesis was something for which I am truly grateful.

I must thank my colleagues, Eleanor Drake, Ian Finlayson, Dr Paul Brocklehurst, Madeleine Beveridge, Ollie Stewart who offered encouragement, expertise and time, you made the PhD journey all the better.

Finally, to my friends and family, I literally could not have done this without you. Special thanks go to my family for their endless support; you have been involved with every step of the process along with me, often without your choice. You helped give me the confidence and determination for such an undertaking in the first place and for that I will be forever thankful.

Contents

List of Figures

List of Tables

CHAPTER 1

1.1 Introduction

In many language studies speakers are considered to be consistent in their delivery, in so much as that they always produce fully formed and fluent speech. Despite high levels of aptitude for speech displayed by humans, it rarely proceeds without deviation from a fluent production. Speakers often repeat or correct parts of the utterance they have already produced or their speech is interrupted with filled pauses, such as *um, uh* and *er*. Disfluency is the term used to describe this range of speech phenomena.

Spoken language comprehension is impacted by the presence of disfluencies. There have been attempts to understand the underlying mechanisms that are responsible for these disfluency effects. This thesis tests the roles of attention and prediction during disfluent language comprehension as a starting place to explore the predictional and attentional accounts of disfluency processing. These accounts of disfluency and the related mechanisms are investigated in this thesis using both eye-tracking and speech perception methodologies.

1.2 Thesis Overview

The aim of the current thesis is to investigate the roles of attentional and predictional mechanisms in disfluency processing, as a window on investigating both the predictional and attentional accounts of disfluency.

In Chapter 2 we discuss the relevant literature relating to the areas of central focus in this thesis. We start by exploring the topic of speech perception, which is tested during the majority of experimental study below (Experiments 2-6). Our aim with this section is demonstrating that it is uncontentious that listeners can vary their use

of bottom-up and top-down processing during perception, as demanded by the context of what is being heard and the task being undertaken. We are especially concerned with providing clear evidence that focused attention can drive listeners to increase their use of bottom-up processing resulting in an increased sensitivity to speech at a phonemic level.

The next topic discussed in Chapter 2 is prediction, which is essential to our understanding of defining a predictional processing mechanism for the investigation of the predictional account of disfluency. We explore the higher level, linguistic information that can be used to generate expectations about upcoming content based on the preceding context that cannot be explained by a purely bottom up processing account.

Following this, the next section is concerned with disfluency, the central phenomenon being explored in the current thesis. Firstly, a discussion of what disfluency entails is provided, before a basic taxonomy of disfluencies. Extra emphasis is given to the variant of disfluency that is employed in the empirical study that follows, filled pauses. We provide in-depth discussion of the stated effects that relate to the current research on disfluency, prediction and attention in language comprehension. We then define both the attentional and predictional accounts of disfluency processing with supporting evidence. These accounts form a central topic of study in this thesis.

The final section within Chapter 2 focuses on the visual world methodology, which is used in Experiment 1. We highlight the validity of visual world studies to inform about the online processing of language and how this methodology has provided insight and evidence for the effects of both disfluency and prediction on the processing of language.

Chapter 3 explores a topic of interest for the thesis that is based around addressing

the question of the role of attention and prediction in disfluency processing using visual world methodology. Experiment 1 investigates the online processing of language with incidence of disfluency using an eye-tracking paradigm, with a view to differentiating between predictional and attentional accounts of disfluency.

In Chapter 4, we change methodology, exploring the attentional account of disfluency processing further by employing a speech perception paradigm. This Chapter contains Experiments 2 and 3, in which we use a sentence based lexical decision task that asks participants to judge a pair of target words that contain word initial phoneme variation to investigate the effect of focused attention on responses. In these experiments we use only a fluent production as we aim test the effect of attention before we complicate the paradigm with the inclusion of disfluency.

In Chapter 5 we continue to address finding an attentional effect with an updated speech perception paradigm. We move the phoneme variation to a word medial location. Again, in this study only a fluent production is used. However, in Chapter 6 we add disfluency into the paradigm used in the previous study, creating a crossed focused attention and disfluency design that allows us to explore the role of each in language comprehension.

In Chapter 7, our final experiment, we update the paradigm to centre on the phoneme variation that produced the largest differences seen in the previous 4 experiments (2-5). We also increase the number of targets being tested, so that any effects would better generalise across target words.

Finally, Chapter 8 provides a summary of the findings from the current thesis and discussion on their possible implications for the mechanisms that underlie disfluency effects and the impact that this has on accounts of disfluency processing during language comprehension.

CHAPTER 2

Literature Review

2.1 Chapter Overview

In the introduction we presented an overview of the thesis and the central themes: disfluency, attention and prediction during comprehension. In the current chapter we review the literature relevant to understanding these central themes, with the aim of showing the motivation for the experimental research that follows. This chapter is broken down into different topics of interest, each forming a section. We start with an overview of speech perception, exploring how contextual information and attention can bias a listeners' perceptual processing. Following this, we introduce the literature on the phenomena of prediction during language comprehension, asking what this process entails and the expectancy effects that have been demonstrated so far. Our discussion of disfluency follows. This is the central phenomenon being investigated. We provide a summary of disfluency and the effects seen in comprehension before we focus on the topic of primary interest, filled pauses, and the impact that they have on listeners. After, we discuss the sensitivity of the visual world methodology to capture linguistic effects pertaining to comprehension, prediction and disfluency. Finally, this chapter ends with the topic of attention and how it is defined for the current thesis. Overall, this chapter aims to demonstrate the current state of the field in relation to our research question about how predictional and attentional accounts of disfluency processing impact upon listeners' comprehension.

2.2 Speech Perception

Listeners are adept at successfully decoding variation in the acoustic speech signal into a coherent and useful message. The interest for the current thesis is in how attention and prediction impact upon the processing of speech perception. Although we are not concerned with modelling speech perception, our focus is on

the underlying mechanisms and the contribution of top-down and bottom-up processing and how they interact when faced with variation, in this case disfluency, in the speech signal. Investigating the distinction between the effects of employing top-down and bottom-up information during perceptual processing is crucial, as it acts as a window on the central theme of the current thesis: the understanding of disfluency during comprehension and its relevance for different theoretical accounts, which diverge on the use of top-down and bottom-up processing following disfluency. Attention has been shown to exert influence on how pronunciation variation is perceived, appearing to focus participants on the bottom-up acoustic information contained in the speech stream. In reviewing the speech perception literature, we aim to demonstrate that there is clear evidence to implicate that focused attention during comprehension leads to the use of more fine grained acoustic detail and less reliance on top-down information. In doing so, we examine what role our expectations, experience of language and attentional state play in how we perceive speech and how this informs the current thesis.

This section intends to answer this question, by, first giving an overview of the contextual information which has been shown to impact upon speech perception. Following this, we outline the consequences of context for modelling speech perception and additionally, how these models integrate top-down versus bottom-up information. Finally, we explore the role of focused attention in choosing between these two sources of information during perceptual processing of ambiguous speech.

2.2.1 Perception of Variation in Speech: Contextual Effects

"Pronunciation variation can be thought of as perceptual adversity for the listener," Pitt & Szostak (p1226: 2012).

Speech production is highly variable, with different speakers displaying variable acoustic characteristics (e.g., Johnson, Ladefoged, & Lindau, 1993) and this leads to

realisation of phonemes varying across speakers (see Peterson & Barney, 1952). This pronunciation variation is widespread with many non-canonical phoneme forms being seen in corpora of speech (see Bell et al., 2003; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005). A central theme of speech perception research is how this variation in the speech signal can be understood by listeners with apparent ease (e.g., Norris, McQueen, & Cutler, 2003). Contextual information at auditory, lexical, semantic and sentence level has been shown to influence how speakers perceive speech.

2.2.2 Acoustic Contextual Effects

The acoustic context that a phoneme occurs within can affect how it is categorised: Identical speech stimuli can be categorised as different phonemes depending on neighbouring phonemes (see Repp, 1982). In an influential exploratory study, Lindblom and Studdert-Kennedy (1967) reported that for synthesised speech-like sounds participants' categorizations of a vowel sound were dependent on the acoustic information of the surrounding environment. They used a forced-choice identification task, in which participants heard a CVC syllable and were then asked to identify the vowel sound from one of two choices. This study suggested that perception was not affected by only the formant frequency of the vowel sound but also by the formant information adjacent to this vowel sound. However, this study did not employ natural speech sounds and asked participants to identify monosyllabic nonsense speech stimuli; this questions the ecological validity of the results in relation to natural speech.

However, it proved a useful starting point to investigate the influence of the auditory environment on phonemic categorisation. Mann (1980) demonstrated that for judgements on a range of speech stimuli from /ga/-/da/, forming a perceptual continuum, participants changed their categorisation based on whether the stimuli was preceded by either /al/ or /ar/: Participants made more /ga/ responses when the stimuli were combined with /al/, whereas the same speech stimuli were perceived more often as /da/ when preceded with /ar/. An effect of differing phonemic

categorisation for a similar series of /ga/-/da/ speech stimuli can even be elicited by using synthesised high and low non-speech tones to represent the different formant frequency characteristics of /al/ and /ar/ in natural speech production (Lotto & Kluender, 1998). This shift effect for categorisation of vowels has been replicated using different consonant and vowel combinations (Holt, Lotto, & Kluender, 2000). Holt (2005) extended this area of study to show that non-speech tones can effect categorisation of non-adjacent speech stimuli: Non-linguistic tones still shifted categorisation after either intervening silence or spectrally neutral acoustic stimuli for over one second before the speech stimuli. Across these non-speech tone studies (Holt, 2005, 2006; Lotto & Kluender, 1998) the authors have theorised that the mechanism that underlies the categorisation effect is a contrastive frequency mechanism that relies on the differences in frequency between tones: either high or low frequency. The high frequency tones elicited more /ga/ judgements from participants in response to a range of speech stimuli, whereas the same stimuli preceded by low frequency tones drew more /da/ responses.

As noted above, there is massive variation in the acoustic characteristics of how speech and phonemes are produced between speakers, due to differences in, and not limited to, vocal tract, gender, age and language variants spoken (see Johnson, Ladefoged, & Lindau, 1993). So how does this talker-specific acoustic variation map on to a contrastive frequency mechanism (e.g., Holt, 2006) to categorize speech when each speaker is likely to have different sets of frequency contrasts? Variable vocal tracts cause systematic differences in acoustic signatures related to the relevant vocal tract, meaning that listeners can make use of these regulated acoustic characteristics to accurately perceive speech. Talker specific categorisation based on systematic frequency changes has been empirically supported: Ladefoged and Broadbent (1957) found that the formant frequency characteristics of the vowels in a sentence context influenced the categorisation of a vowel in a constant target word (/b_t/). Certain formant frequencies were shifted up or down in the vowels of the sentence context and this systematic change corresponded to a change in vocal tract

of a speaker. The raising of the formant in the preceding vowels led the target vowel to be perceived as 'bit' more often; whereas the lower frequency formant in the preceding vowels led to 'bet' being heard more often. This finding supports the belief that listeners show sensitivity to talker specific categorisation.

More recently Laing, Liu, Lotto, and Holt (2012) have proposed that the crucial characteristic taken from auditory context which underlies talker specific speech categorisation is the long-term average spectrum (LTAS) of a speaker. Adapting the materials from Ladefoged and Broadbent (1957) they manipulated sentence contexts to sound like different speakers: they manipulated two different sets of formants within the vowel (F1 and F3). Listeners then had to categorise a target stimulus (/ga/ or /da/). Only one set of formant manipulations (F3) resulted in a categorisation change for the target: the F3 has acoustic energy relevant to the /ga/-/da/ distinction, whereas the F1 variation does not. This demonstrates that the categorisation difference is down to task relevant auditory characteristics. These studies provide evidence that speech perception is sensitive to a range of neighbourhood acoustic information (Ladefoged & Broadbent, 1957; Laing et al., 2012; Lotto & Kluender, 1998) and that these acoustic characteristics do not have to be adjacent to a target stimulus to affect speech perception(Holt, 2005).

It is not just the acoustic information carried in speech contexts that can affect listeners' perception of speech sounds; temporal cues can also cause variation in how speech stimuli are categorised. Phoneme distinctions can be shifted by a change in duration of neighbouring acoustic signal (see Diehl, Lotto, & Holt, 2004). Aside from purely acoustic frequency or duration distinctions influencing speech perception, listeners undergo a perceptual learning process where they map variation in a speaker's pronunciation of a phoneme onto the intended matching phonemic representation. The perceptual systems are flexible when it comes to variation in speech signal and even a small amount of input from a new speaker causes a listener to rapidly adapt to that speaker's pronunciation and adjust phoneme distinctions to be able to comprehend that speaker and improve

subsequent instances (Kraljic, Brennan, & Samuel, 2008a; Kraljic, Samuel, & Brennan, 2008b; Kraljic & Samuel, 2011). In Kraljic et al. (2008a) listeners showed altered perceptual sensitivity to an acoustically identical ambiguous phoneme, midway on the /s/ to/ʃ/ continuum, based on the contextual setting of the different sources of the speech segment; Participants heard the variation as resultant from either an idiolectal context, where all realisations of this phoneme were the same, or dialectal context, where the phoneme was only realised in this ambiguous manner in particular phonetic contexts. The results showed that participants treated the variation differently based on the source, with a perceptual learning effect seen for the pronunciation variation linked to an idiolectal context but not for the dialectal equivalent. These results were supported by Kraljic et al. (2008b) who found that pragmatic information, such as whether participants had access to visual information showing that the speaker had a pen in their mouth, impacted upon how the same ambiguous phoneme, midway on the /s/ to/ʃ/ continuum, was perceived. When participants had only audio information to inform their impression of the speaker they accommodated the ambiguous phoneme as a feature of that individual's speech with perceptual learning but when they could see an external reason why the speaker produced the phoneme in an ambiguous manner, no perceptual learning took place. Kraljic & Samuel 2011 extended the understanding of what can information sources may influence the perceptual learning effects seen in the previous two studies by Kraljic and colleagues, showing that visual information can influence perceptual learning but that this is not always the case.

In summary, speech perception is sensitive to the fine grained acoustic characteristics (Laing et al., 2012) time course (Diehl et al., 2004) and speaker specific information (Kraljic, Brennan, et al., 2008a; Kraljic, Samuel, et al., 2008b; Kraljic & Samuel, 2011) heard in the surrounding context. In relation to the current thesis, these factors must be controlled for during speech perception tasks.

2.2.3 Lexical Influences on Speech Perception

A key area of interest for the thesis is the use of top-down processing influencing the perception of the incoming speech signal. How lexical influences, contextual knowledge, and the associated effects impact on a listener's perception of ambiguous speech has been the focus of considerable study. Phonemes are recognised quicker in words than non-words (Cutler, Mehler, Norris, & Segui, 1987) and in more word-like non-words compared to less word-like non-words (Connine, Titone, Deelman, & Blasko, 1997). A topic of particular interest is whether listeners exhibit a bias to perceive ambiguous phonemes within words/pseudowords in ways that are consistent with lexical entities or 'words'.

The phoneme restoration phenomenon provides evidence for lexical bias and that this influence increases as a word unfolds. This restoration effect refer to listeners claiming to have heard a phoneme that was removed, due to lexical influences. In the classic study, Samuel (1981) showed that listeners are able to successfully map between an incoming acoustic signal and a lexical representation, even when there is a phoneme missing or it has been replaced with another noise, perceiving the word as intact. Participants heard a word that had either been acoustically altered to remove a certain phoneme or insert noise over the top of this phoneme. The participants were then asked whether the speech they heard was intact or not. The missing phoneme was restored in a higher proportion of times when it featured in a word context, compared to when it featured in a pseudoword. This reinforces the finding that speech perception is enhanced when listeners have expectations of a lexical entity, driving top-down processing. Samuel (1987) then employed the phoneme restoration effect to investigate the strength of lexical influences in different phoneme locations across a word; finding an increasing rate of restoration at the ends of words compared to initially and medially. These results provide evidence of instances of lexical influence and how it grows in strength as words unfold.

In a seminal study, Ganong (1980) tested phonemic identification using a phonetic continuum, that created a word to non-word continuum, when presented in a syllable context. Ganong investigated a voicing continuum that ranged from /t/ to /d/. The target phoneme was presented in a word initial phoneme location creating a continuum of stimuli that ran from either 'dice-tice' or 'dype-type'. This created a word only at only one end of the voicing continuum. Ganong demonstrated that for perceptually ambiguous stimuli, participants tended to categorise them as the phoneme that would result in creating an existing word using the syllable context. This meant that in the _ice context the categorisation boundary shifted so there was a tendency for the phoneme to be identified as /d/, whereas in the _ype context the boundary shifted so that it tended to be labelled as a /t/. This paradigm has been replicated numerous times since, with a reported effect of lexical bias, resulting in a shift in categorisation boundaries (e.g., Mattys & Wiget, 2011; Pitt & Szostak, 2012; Pitt, 2009).

A lexical bias effect has been replicated across a range of phoneme positions within a word context: McQueen (1991) demonstrated a 'Ganong' effect in a word-final location with a fricative continuum ('Harsh-Harce' and 'Presh-Press'). Crucially, in this study a lexical bias in phoneme categorisation only appeared once the speech signal was degraded by employing a low-pass filter. A lexical bias in word final phoneme positions is afforded support by evidence demonstrating mispronunciation/phonemic variation is harder to detect in a word final location (Cole, Jakimik, & Cooper, 1978; Marslen-Wilson & Welsh, 1978). Cole and colleagues found that mispronunciation was detected 72% of the time for word initial phoneme variation in a monosyllabic word, compared to only 33% of the time when the phoneme variation occurred word finally.

Mirman, McClelland and Holt (2005) showed that a lexical bias effect in speech perception can extend to phonemes that are similar to an expected phoneme within a word. They tested listeners' sensitivity to lexical influence by using three classes of stimuli: words (W), near non-word (NNW) and distant non-words (DNW). To

create the non-words they replaced a target phoneme (e.g., /s/) within a word (e.g., goddess) with one of two variants: One phoneme variant that was similar to the target phoneme (NNW) (e.g., a change from /s/ to /ʃ/) and one that is not similar (DNW) (e.g., a change from /s/ to /k/). The similar phoneme is equivalent to the non-word end in a Ganong paradigm. They found that listeners were slower to detect a target phoneme in a NNW than a DNW. They suggest that due to the similarity between the replaced phoneme in the NNW and the target phoneme there is some top-down lexical influence that causes a delay in phoneme recognition, when compared to the DNW where there is limited influence of lexical bias due to the replacement phoneme used. This finding supports a lexical bias effect, as detailed previously, and suggests that both top-down and bottom-up processes are attended to during speech perception. If there was no lexical bias effect (a top-down process) the bottom up acoustic information heard by listeners would create non-words for both variants of replacement phoneme at the same temporal point. However, that is not what is seen, as there was a delay for the NNW containing the similar phoneme. In addition to lexical bias, listeners have experience of perceiving some words more than others and this affects their perception: At a lexical level, listeners' phonemic categorisation was impacted by the frequency of a lexical entity (see Pinnow & Connine, 2013). Familiarity with a frequent surface form of a word, whose pronunciation may include variation such as the partial or full deletion of a phoneme, leads to easier categorisation of these forms compared to low-frequency forms (Connine, Ranbom, & Patterson, 2008). Connine et al. (2008) showed quicker lexical decisions for more frequent surface forms of words compared to less frequent surface forms.

2.2.4 Sentential Context on Speech Perception

In the current thesis, we are interested in the comprehension of disfluency and how it intersects with prediction and attention at a sentence level. Therefore, it follows that there is a necessity to understand the effect of context on speech perception processes at sentential level. There are clear perceptual findings at a lexical level but

how does context extend to affect speech perception at the sentential level? We have already detailed how changes in the acoustic properties of sentence contexts can impact the categorisation of a target speech segment (Ladefoged & Broadbent, 1957; Laing et al., 2012). The semantic information carried by a sentence can also create expectations that impact categorisation; Connine (1987) investigated the effect of semantically biased sentence contexts on target words whose initial phoneme varied along a voicing continuum (e.g., 'tent' to 'dent'). Subjects had to press a button (either 'T' or 'D' for the example) to categorise the phoneme heard. She found that perceptually ambiguous phonemes were labelled to create a word that fitted semantically with the sentence context. Borsky, Tuller, & Shapiro (1998) replicated this finding with the addition of a visual probe. The target word varied along a word initial voicing continuum from 'coat' to 'goat'. Participants heard sentence contexts which biased towards one of the interpretations by employing selective verbs, such as 'milk' or 'dry clean'. After hearing targets with ambiguous phonemes, a visual probe word was presented, either 'goat' or 'coat'. Participants had to judge whether the stimulus heard matched this probe or not. Participants' judgements were biased by the verb heard; after hearing 'milk' they chose 'goat' more often and following 'dry clean' they chose 'coat'.

It is not just semantic content that can be influential at a sentential level. Pragmatic cues in the preceding sentence context can impacted how a phoneme is categorised (Rohde & Ettlinger, 2012). Syntactic structure has also been shown to influence the categorisation of a /t/ phoneme in Dutch (Tuinman, Mitterer, & Cutler, 2013). In Dutch, speakers have experience with the deletion of word final /t/, with tokens ranging from present to fully deleted. Although naturally occurring tokens of forms of /t/ deletion exist, a synthesised continuum was created for accuracy, running between these endpoints, with ambiguous forms in the middle. The results showed that participants categorised the ambiguous phonemes as present in more cases in sentence where the presence of /t/ would be syntactically correct, in comparison to sentences where it would not be.

The temporal unfolding of sentential context effects is another factor that has been investigated as to the impact it has on speech perception. When listeners were put under time pressure and had to make speeded responses this significantly decreased the impact of lexical and sentential contexts (Miller, Green, & Schermer, 1984; Miller & Dexter, 1988). This suggests that differences in temporal pressure cause variation in listeners' integration of lexical and sentential contextual information.

Borsky, Shapiro and Tuller (2000) further tested the influence of time-course and sentence context on speech perception using cross modal interference and word monitoring. As in previous Ganong style studies (e.g., Ganong, 1980), they employed sentences biasing towards one endpoint of a 'goat' to 'coat' continuum (Borsky, Tuller, & Shapiro, 1998). In the cross modal interference task, participants' primary aim was to listen to these sentences. Participants were then presented with a visual probe, either a word or orthographically possible non-word, at different points throughout the sentence, in relation to the target words: control, immediate and delayed. The control probe was before the target was encountered. The immediate probe was presented at the offset of the target word. Finally, the delayed probe was not presented until 450ms after the target word. Participants had to make a lexical decision about the probe. The time course of the probe presentation, relative to the target, affected the response times of participants: when the probe was presented in the immediate position there was no difference in response times by contextual bias of the sentence; in the delayed position, the context of the sentence had a clear effect, with significantly longer response times. This suggests that there was a lack of sensitivity to the sentence context early on in the perceptual processing of the target word. In the word monitoring task, participants had to respond to a target word in a sentence by pressing a button as quickly as they could. The results for this experiment showed that there were significantly longer response times for target words that mismatched with the sentence context. Taken together, these findings are consistent with the influence of semantic information from a

sentence context on speech perception but this effect was modulated by time course and task.

2.2.5 Speech Perception and Cognitive Load

Speech is not always confined to a laboratory and in naturalistic settings there is often background noise or competing sounds that degrade the speech signal, yet, still listeners can comprehend the speech signal in a competent and efficient manner. However, this additional auditory information does impact upon speech perception (for a review see Mattys, Davis, Bradlow, & Scott, 2012). Similarly, often during speech there are other expectations on cognitive processing and the effect of cognitive load and its impact on speech perception has been investigated: Mattys and Wiget (2011) tested the strength of lexical bias effect using a Ganong paradigm, whilst comparing the results when listeners were subject to either an additional cognitive load or no load. In the load condition listeners showed an increased use of lexical bias in phoneme identification and decreased reliance on the acoustic information. This points towards that when experiencing high cognitive load listeners exhibit an increased reliance on top-down processing, resulting in greater lexical bias to improve the probability of a successful perception of the target word.

2.2.6 Modelling Speech Perception

As noted in the overview of speech perception, modelling speech perception is not an issue we focus on with the current research. Instead we are interested in the role that top-down and bottom-up processing take in explaining perception of speech variation and whether this can vary as a function of focused attention. However, for an efficient discussion of the contribution of contextual influences and how top-down and bottom-up processes impact perceptual processes, it seems reasonable to provide an overview of the current state of the literature in terms of modelling speech perception.

We have demonstrated that there are multiple information sources integrated during speech perception and phoneme categorisation in particular; there has been

considerable empirical work to try and build a useful model that can account for the variety of effects stemming from auditory, temporal and linguistic contextual information but, as it stands, at the core of the debate are two theoretical standpoints: inference-based (see Gaskell & Marslen-Wilson, 1998) and episodic/representational (see Goldinger, 1999; Ranbom & Connine, 2007). Gaskell & Marslen-Wilson (1998) and the inference based accounts theorise that recognition of pronunciation variation within words is undertaken by matching them to a single, abstract lexical representation held in lexical memory. Inferential processing transforms the mismatching acoustic information to the corresponding lexical representation. Episodic models (e.g., Goldinger, 1999), differ by having multiple representations with a separate representation for each of the different ways a word maybe realised. These variant representations are all linked to a single lexical entry. So pronunciation variation is perceived by matching incoming auditory information against one of these lexical representations held in memory. Representational models (e.g., Ranbom & Connine, 2007), diverge from episodic models in the way that lexical representations are selected. Currently, a third 'hybrid' account (see Pinnow & Connine, 2013; Pitt, 2009) has been proposed in response to neither inferential or episodic models being fully able to explain the variability of speech. Inference alone cannot fully account for how novel variants are recognised as known words. If a novel variant of speech occurs in a new phonological environment, then no rule should exist to be able to infer the connection to a previous lexical representation and the listener would not be able to comprehend the speech. However, listeners can adapt to variance in the pronunciation of speech even in novel contexts or with idiolectal variation (e.g., Kraljic, Brennan, et al., 2008; Kraljic, Samuel, et al., 2008). Additionally, there are clear advantages in having a stored representation to access for regular variants and new variants that become generalised with experience of their form. These representations extend beyond the proposed episodic accounts (see Pitt, 2009; Sumner & Samuel, 2005). A hybrid accounts is able to incorporate the nuances gained from greater linguistic experience, leading to an increase in recognition of variable pronunciations of

words. Although there is still ongoing debate about the details of these accounts, they all support higher level representations facilitating top-down processing from contextual information.

2.2.7 Contextual effects in modelling Speech perception: the integration of top-down versus bottom-up information.

Following on from our general overview of speech perception models, we focus in on the debate surrounding the integration of top-down versus bottom-up information in modelling to meet our stated aim of understanding the role that top-down and bottom-up processing take in the perception of variation in speech and the effect of focused attention. Bottom-up processing attends to the fine acoustic detail of incoming speech signal and builds up representations from the input, sending the selected representation to a higher level until a word is perceived.

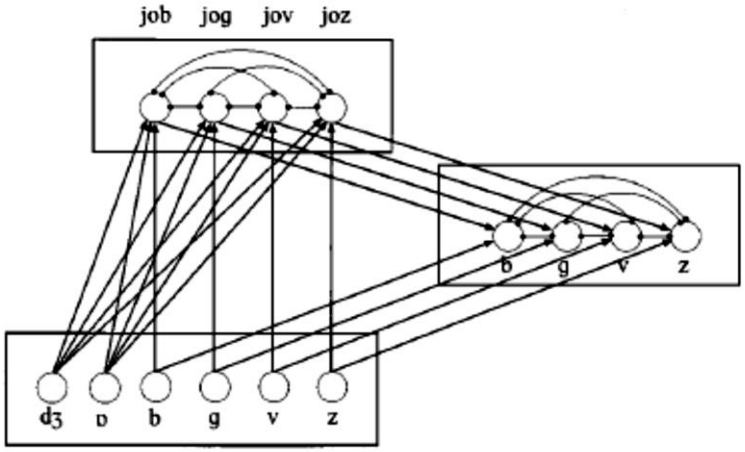


Figure 2.1- Merge Model- The basic architecture of the merge model. Excitatory connections are shown as bold black lines with arrows; activation can spread only the way of the arrow. Inhibitory connections are shown with fine lines and closed circles; activation can spread both ways. From the bottom, the input nodes spread activation to the lexical level (top) and phoneme-decision level (right). Figure taken from (Norris, McQueen, & Cutler, 2000).

The processing advantages of the inclusion of top-down processing are easier recognition of speech sounds; if higher level lexical representations become

activated during the perceptual process then they provide a secondary source of information to activate the phonemes employed in a word representation, requiring less acoustic information for the phonemes to be recognised than without this lexical level information. First, we present a summary of the Merge model (Norris, McQueen, & Cutler, 2000), as an exemplar of a speech perception model to elucidate the typical levels thought to be involved with perceptual processing and to provide context for the debate about the realisation of top-down versus bottom-up processing.

The Merge model features 3 levels, see Figure 1: Pre-lexical, phoneme-decision and lexical. So to run through a typical perception process, participants first hear a string of acoustic information, for example, "job" in the figure above. The first action is pre-lexical processing that activates the relevant input nodes: /dʒ/, /ɒ/, /b/. This pre-lexical processing then provides continuous information to the lexical level in a bottom up manner. This activates compatible lexical candidates. The pre-lexical processing, simultaneously, provides information to the phonemic decision level, allowing compatible phonemes to be activated. The post-perceptual phoneme decision stage can merge the information from both the lexical and phonemic-decision levels, providing lexical influences on phoneme decision, therefore providing the model with a mechanism for top-down processing, responsible for contextual effects such as lexical bias in Ganong tasks (Ganong, 1980; McQueen, 1991; Pitt & Szostak, 2012), phoneme restoration tasks (Samuel, 1981; 1987) and sentence level semantic effects (Borsky et al., 1998; Connine, 1987).

However, there is currently fierce debate over the nature of processing during speech perception, focusing on how top-down processing interacts with lower levels in speech perception models. Proponents of interactive models (e.g., TRACE, Mcclelland & Elman, 1986) argue that top-down processing can influence speech perception directly, with lexical bias exerting control over the lower level of phonemic analysis (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland &

Elman, 1986; McClelland, Mirman, & Holt, 2006; Samuel, 2001). However, there has been criticism of the interpretation of some sentence level effects in favour of interactivity (e.g., Miller et al., 1984); it has been argued that the contextual effects shown may stem from the frequency of co-occurrence between words occurring in close proximity (e.g., 'water'-'bath'), rather than due to top-down semantic processing. For supporters of opposing 'autonomous' models (e.g., Merge, Norris, McQueen, & Cutler, 2000), these top-down contextual effects are theorised as occurring at a post-perceptual decision stage, as described above (McQueen, 2003; McQueen, Jesse, & Norris, 2009; Norris et al., 2003).

Further, research claimed to demonstrate lexical feedback that could not be accounted for by frequency effects: Magnuson et al. (2003) investigated lexically-mediated compensation for co-articulation during speech perception. This phenomena describes how lexical knowledge responsible for categorising an ambiguous sound (e.g., That when a final ambiguous fricative is used in 'Christma?' this phoneme should be categorised as /s/) also affects the perception of the following sound. Their results demonstrated that there was still a lexical influence on word recognition, even when the frequency effects went against the lexical bias. However, these findings have since been challenged, with the lexical influence determined to have been caused by statistical biases from the pre-test practice block influencing the transitional probabilities of coarticulation in the main experiment (McQueen et al., 2009).

More recent studies aiming to weigh in on how contextual effects are modelled in speech perception have been relatively ambiguous: The study of pragmatic cues provided divergent evidence across experiments, leading Rohde and Ettlinger to suggest that the variance in the contextual effects seen may require different approaches towards top-down processing to account for them (2012). Tuinman et al. (2013) suggest that the variable patterns seen are resultant from the nature of the task: listeners relied more or less on the contextual information as a function of the

other information available to them that can help solve ambiguity in the speech stream. Tuinman and colleagues showed that in the face of an ambiguous sound the syntactic context of a sentence is used to inform the listeners' decision about whether or not a phoneme is present. They leant towards a post-perceptual explanation as it better reflected the different patterns seen for the syntactic context effects that they found across experiments, supporting similar results found by Van Alphen and McQueen (2001). The debate continues but what is becoming apparent, as Tuinman and colleagues make clear, is that there can be a degree of listener control exerted over the acoustic information attended to during speech perception and that this is based on task demands.

2.2.8 Speech Perception and Attention: the variable role of top-down and bottom-up processing.

An area of particular interest for the current thesis is the intersection of speech perception and attention. There is a wealth of evidence that listeners' speech perception abilities can be variable due to differing task demands: cognitive load (Mattys & Wiget, 2011); time course (Miller et al., 1984; Miller & Dexter, 1988); lexical bias (Ganong, 1980; Pitt & Szostak, 2012; Pitt, 2009; Samuel, 1987); pragmatic cues (Rohde & Ettlinger, 2012) and syntactic context (Tuinman et al., 2013). There have been a number of demonstrations of the effect of attention on speech perception. Cutler et al. (1987) implicitly showed that a change in participants' attention to incoming speech signal can impact how they perceive that speech. Participants had to listen to word pairs and press a button when a specified target phoneme was heard (e.g., /d/) and their reaction times were measured. The target phonemes were housed in either words or non-words and were presented in one of two conditions: repeated or variable stimuli. The word stimuli were predicted to have a reaction time advantage due to top down lexical effects. A lexicality effect did appear when there was variation in the targets. However, their findings revealed that these lexical effects disappeared in the repeated condition. Cutler et al. (1987) theorised that the cause of this variation in lexical influences was due to a

shift of attention caused by repetition; in the repeated condition, after hearing a target numerous times participants pay less attention to the semantic content of the stimuli and increase their focus on the detection task. This causes them to attend to the incoming acoustic information more and is reflected in their decreased reaction times. It can be argued that the phoneme detection task employed does likely not reflect the comprehension processes in normal speech perception due to the necessity of attending to this fine grained acoustic detail. However, the study reveals how task demands can implicitly change participants' locus of processing, reflected in participants' variable attention to either top-down or bottom-up processing in speech perception. Further support for the dynamic role of perceptual processing in speech stemming from task demands has been evidenced in a word identification task (e.g., Miller et al., 1984). Miller et al. reported that changing the task demands so that listeners' focus was directed to only the target word and away from the surrounding sentence context caused the disappearance of a context effect. This again suggests that when listeners change strategy in a task it can affect how they attend to contextual information, leading to a greater reliance on bottom-up processing.

In an explicit demonstration of attention in the perceptual processing of speech, the phoneme restoration effect was nullified when participants were instructed as to the phoneme location that is impaired or removed (Samuel & Ressler, 1986). Attentional and lexical influences were manipulated between participants using 4 variants of visual prime: i) a control group, where the prime did not contain any information about the test word; ii) a group where the prime cued the upcoming test word, whilst marking both the location and identity of the critical phoneme using an asterisk; iii) a group where the prime cued only the upcoming test word but not the critical location or identity of the critical phoneme and iv) a group where the prime cued the identity and location of the critical phoneme using an asterisk but not the test word. The attentional manipulation was the signposting of the identity and location of critical phoneme; whereas the lexical manipulation was whether the

participants saw the test word or not. The results showed that an attentional effect was only present when both the attentional and lexical manipulations were known to the listener prior to the test word (Condition ii). Signposting only the identity and location of the critical phoneme without having prior knowledge of the test word (Condition iv) resulted in a slight decrease in performance in comparison to the control group. The findings here again implicate sensitivity to an attentional mechanism in the speech perception process, with a replication of a decrease in lexical influence when attention is cued towards the incoming auditory information.

Mirman, McClelland, Holt and Magnuson (2008) are proponents of an attentional mechanism being added to interactive models of speech perception (e.g., TRACE, McClelland & Elman, 1986) to account for the variable perceptual processes that occur due to either implicit task demands or explicit focused attention. In two phoneme detection studies they found that when attention is focused on incoming speech this attenuates the activation from lexical influences. Their findings are strengthened by running simulations using their updated TRACE model, with the results following the outcome of the behavioural studies. From a processing perspective, Mirman et al. suggest that focused attention modulates participants' reliance on top-down contextual processing, in this case limiting lexical activation, in favour of the bottom-up integration of fine acoustic detail during the perception of speech sounds.

Since Mirman et al. have proposed this, there has been further evidence to support their inclusion of an attentional modulation in speech perception. Pitt and Szostak (2012) found that by varying the attentional focus of listeners they could induce changes in the proportion of stimuli that were categorised as lexical. The explicit attentional manipulation employed was the instructions that participants saw; either cueing the critical location and the identity of the target phoneme within a target word or not. Using a Ganong style paradigm with a 5-step /s/ to /ʃ/ phoneme continuum, they also investigated the attentional and lexical influences across phoneme locations with variation occurring in word initial, medial and final

locations. They chose target word pairs for each phoneme location (e.g., word intial: 'serenade' and 'chandelier') that were lexical at one of the continuum and a non-word at the opposing end ('sherenade' and 'sandelier'). They asked listeners to complete a lexical decision task, either labelling the target word that they heard as a 'word' or 'non-word'. A lexical bias was seen in the unfocused condition with participants showing a tendency to label ambiguous stimuli in line with lexical influences. However, in the focused attention condition participants labelled a lower proportion of target stimuli as 'words'. Larger differences between attentional conditions were seen in the word medial and final phoneme locations compared to the word initial location. The findings here replicate an attentional effect and provide further evidence that participants can exert control of how they attend to contextual and fine grained acoustic information during speech perception. Pitt and Szostak concur with Mirman et al.'s (2008) proposal that attention acts to damp lexical influences, further supported by the increasing effects of attention across word positions. As Pitt and Szostak state, lexical influences should increase with the more of a word that is heard, which are exactly the results seen.

Taken together, these studies provide clear evidence for the influence of an attentional modulation in speech perception; by using either task demands or an attentional manipulation, listeners' perception of speech can be impacted to a lesser or greater extent by the integration of bottom-up or top-down processing.

2.3 Prediction

In the current thesis we are concerned with contrasting predictive and attentional accounts for the comprehension of disfluency. Therefore, it is crucial to show that prediction exists and cannot be explained by an integrative account. Firstly, we will define what prediction entails and then outline a range of effects seen previously for anticipatory processing and then detail how these could impact our study.

Prediction, as it is understood here, relates to the use of higher level, linguistic information to generate expectations about upcoming content based on the preceding context that cannot be explained by a purely bottom up processing account. The higher level information can be from a number of linguistic categories such as semantic and syntactic level information. The use of predictive processing is not unique to language, applying to many areas of perception (see Bar, 2009). Expectation based probabilistic models of language comprehension have examined the idea of prediction in comprehension with results that account for much of the behavioural data (e.g., Levy, 2008). Another account links the prediction effects seen to comprehenders simulating the language production of the speaker (Pickering & Garrod, 2007, 2013) However, we are not concerned with explaining the modelling of underlying processing of prediction for the current thesis. Here we detail how people predict other people's language (see Federmeier, 2007; Pickering & Garrod, 2007 for reviews).

2.3.1 Prediction and Semantics

The most attested predictive effect from behavioural paradigms is the semantic context of a preceding sentence exerting influence on a sentence final target word (e.g., Schwanenflugel & Shoben, 1985). In Schwanenflugel and Shoben (1985), listeners showed facilitation effects for the comprehension of highly predictable words from the preceding semantic context. For example, participants heard 'John kept his gym clothes in a...' and participants were faster to process the expected ending of 'locker' over the less predictable 'closet'. This facilitation can be attenuated by the type of word used, with abstract words more sensitive to the preceding semantic information than concrete words (Schwanenflugel, Harnishfeger, & Stowe, 1988). Although the facilitation effects here could be due to easier integration of the word into the surrounding linguistic context, more recent empirical work provides evidence that is harder to account for from this theoretical standpoint, these two approaches are discussed below in 2.3.3. Additional evidence that supports semantically based prediction effects is found in studies based on eye-tracking

paradigms (e.g., Altmann & Kamide, 1999). The relevant findings originating from this literature are discussed below in the visual world section.

Further prediction effects resulting from the semantic information contained in a sentence context have also been demonstrated using ERP studies. The N400 has become known as a measure of the processing associated with the semantic information from speech during comprehension (see Kutas & Federmeier, 2000). The amplitude of the N400 ERP component reflects the semantic fit of a word with the preceding context (Kutas & Hillyard, 1980). In a seminal study, Kutas and Hillyard (1984) found that when a semantically unexpected noun, such as 'coffee' was heard following a biasing sentence context, 'He liked lemon and sugar in his..', it resulted in an increased N400 effect compared to a highly predictable noun, such as 'tea'. This N400 effect has been replicated whilst controlling for different patterns of variation in the sentential context and target word. For example, it has been shown that there is a reduced N400 effect for anomalous words semantically related to a predicted word for the preceding context (Federmeier & Kutas, 1999). In Kutas and Hillyard (1984) the strength of contextual constraint between a sentence and a target word was measured using probability of completion from a cloze paradigm (Taylor, 1953). As we did the same, the use of this measure is of particular interest.

In a more recent study Federmeier, Wlotko, De Ochoa-Dewald, and Kutas (2007) employed a high and low cloze distinction to explore the impact of both strongly and weakly constraining sentences on expected and unexpected sentence final words. Cloze probability reflects the strength of a sentence constraint: a high cloze probability necessarily predicts a highly constraining sentence context. However, both strongly and weakly constraining sentence contexts can be completed with a low-cloze word; that is to say a word that is plausible in the context but not expected. The results of this study replicated the graded nature of the N400 effects seen in Kutas and Hillyard (1984): expected words elicited a smaller N400 component across both strengths of sentence constraint compared to unexpected

words and there was a reduction in amplitude for the N400 component for the expected words in the strongly constraining sentences compared to the weakly constraining contexts. The unexpected words showed comparable N400 components in both strengths of sentence contexts. These N400 effects support a graded constraining effect of sentence context on processing of a word: the degree to which a word matches the information provided by preceding sentence context affects the amount of processing needed to comprehend it. However, there was an additional late occurring (500-900ms) component (P2) seen for unexpected words after a strongly constraining sentence. This effect was only seen following a highly constraining sentence suggesting that the prediction of a specific upcoming word and its associated semantic features exerts a processing cost at a later occurring, temporally distinct period. A processing component of this nature would demonstrate multiple predictive processes occurring at different linguistic levels within comprehension. Taken together these studies show growing evidence for the online anticipation of upcoming words based on the semantic information of the sentence context that precedes it.

2.3.2 Prediction and Syntax

Prediction does not apply exclusively in a semantic domain. There have been demonstrations of predictive processing based on the syntactic information included in a context. In an influential study, Van Berkum, Brown, Zwitserlood, Kooijman and Hagoort (2005) found that Dutch listeners were sensitive to a grammatical gender mismatch between an adjective and an upcoming predictable noun. Participants' ERP activity was measured whilst they listened to a short discourse, for example "The burglar had no trouble locating the secret family safe. Of course, it was situated behind a.." that gave rise to an expected noun completion, for the example "painting". Preceding the sentence final noun, an adjective occurred, either matching or mismatching the grammatical gender of the predicted noun. Their findings revealed an ERP effect locked to the adjective in the mismatching condition, which showed participants' sensitivity to the unfolding syntactic

violations between the adjective and their predictions.

Wicha, Moreno and Kutas (2004) showed syntactic prediction in Spanish. In this study, participants read word-by-word a range of medium to high constraining sentences whilst ERPs were measured to an article and noun within these sentences. The target noun either fitted with the semantic meaning of the unfolding sentence or not. The noun also created a match or mismatch for grammatical gender with the preceding article. The article itself was also expected or unexpected from the sentence context. The results showed that there was definite ERP components related to each of the possible mismatches: An N400 was seen when the nouns did not match the semantic fit of the sentence context; Nouns that elicited a gender mismatch with the article created a P600 effect and finally, unexpected articles drew a larger positive component that occurred 500-700ms post-onset than that seen for the expected articles. These results support readers making predictions about the grammatical category of upcoming words based on anticipated syntax and there was clear effects when these predictions were violated. There is growing additional evidence for the use of prediction associated with the syntactic context of a sentence during language comprehension (e.g., Lau, Stroud, Plesch, & Phillips, 2006; Yoshida, Dickey, & Sturt, 2013). These studies further suggest that in linguistic contexts people make online predictions about specific upcoming words and are sensitive to incremental syntactic information encountered before these words which does not match their expectancies. These instances of anticipatory processing in both syntactic and semantic domains show prediction occurs at differing linguistic levels.

2.3.3 Prediction or Integration?

We have been categorising these behavioural and electrophysiological effects as evidencing the recruitment of predictive processing during language comprehension but there has also been proponents of the viewpoint that some facilitation effects (e.g., Schwanenflugel & Shoben, 1985) could be explained by an

integrative account of language comprehension (e.g., Marslen-Wilson & Welsh, 1978). A predictional account states that language users are creating an expectation of a word from the preceding context; this prediction can then drive a facilitation effect if the expectation is met (e.g., Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1985) or shows a mismatch negativity when an expectation is violated (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). An integration account argues that the facilitation or mismatch effects are down to the level of difficulty in integrating a word's semantic content into a sentence context and the level of resources this requires. Therefore, for a mismatch or unexpected word occurring in a biasing sentence context this would require the listener to devote more resources to process the word than for a highly predictable word where less resources would be required (e.g., Marslen-Wilson, 1989).

However, there is evidence that participants can predict a specific upcoming form and this does not fit well into an integrative account. Readers demonstrate sensitivity to a mismatch effect at the point where their expectation of a particular phoneme and the realisation of a different phoneme diverge (DeLong, Urbach, & Kutas, 2005). DeLong and colleagues exploited the systematically different usage of English indefinite articles that is dependent on the following phonological environment, ('a' before a noun beginning with a consonant sound and 'an' before a noun beginning with a vowel sound). They measured ERPs as participants read sentence contexts, such as 'The day was breezy so the boy went outside to fly ...' The sentence could be finished with either a predictable ending, such as 'a kite' or a less predictable 'an aeroplane. They found a larger N400 effect for the less predictable ending. This N400 effect was graded and correlated with the probability that a predicted word would complete the sentence. However, this effect also occurred on the indefinite article, 'a' or 'an'. This finding supported the pre-activation of certain phonological forms, as the effect started on the preceding article, prior to the phonological realisation of the noun. There is no difference between the articles at a semantic and syntactic level, so can only be the realisation of the upcoming

phonological form. This provides evidence for an effect of predicative processing for the upcoming content, as the effect occurred before the noun, therefore it cannot be down to the ease of integration. This supports what has been evidenced in the syntactic domain (e.g., Van Berkum et al., 2005) as described above.

Another demonstration that is hard to reconcile with an integrative account due to the prediction of the form of an upcoming word is from Dikker & Pylkkanen (2011). It followed on from earlier work by Dikker and colleagues that found early occuring MEG sensitivity to syntactic manipulations in the form of an M100 effect in the visual cortex when an encountered word did not match the predicted syntactic category (Dikker, Rabagliati, Farmer, & Pylkkänen, 2010; Dikker, Rabagliati, & Pylkkänen, 2009). These effects represent the visual properties of the syntactic categories, hence, why the component is seen in the visual cortex. In Dikker et al. (2010) syntactic prediction was controlled using a sentence where participants read either an adjective or adverb preceding a critical noun that either matched or mismatched with the expected word category. Following an adjective (e.g., the beautiful...) a noun would be anticipated, whereas, following an adverb (e.g., the beautifully...) a participle would expected instead. By varying the form of the noun (based on a typicality score) to either look like a typical noun (e.g., soda) or not (e.g., princess) they found that there was an M100 effect when the word read and the predicted category do not match (e.g., the beautifully princess).

In addition an M100 effect has also been shown to be sensitive to semantic predictions of upcoming content. Dikker & Pylkkanen (2011) set out to investigate whether contextual semantic expectation could trigger a similarly early occuring effect in the visual cortex, as seen with syntactic prediction. Participants saw a picture followed by a noun phrase that either matched exactly to the picture seen (e.g., apple) or a mismatch related semantic entity (e.g., A banana). In the mismatch condition there was no expectation generated between the unmatching picture and the noun phrase. There was an M100 effect when there was an exact match between picture and noun phrase but not for the mismatching picture. A third condition was

created where no specific expectation of a unique form could be made due to the picture not representing a single entity but a whole semantic field (e.g., any animal) but the word heard either matched or mismatched this semantic field. For this condition, no differences between M100 component could be seen as a result of a match or mismatch between the word heard and semantic field represented in the picture. These findings suggests that participants were predicting the exact form of the upcoming word, not just the semantic features or a number of forms stemming from a semantic field. These expectations were based on semantic context and produced an M100 effect as seen in the previous demonstrations of syntax based effects. This implies that early occurring visual effects based on prediction of upcoming words are sensitive to information at different linguistic levels.

Overall, it is not contentious to state that people make predictive use of higher-level processing during language comprehension to aid in processing the bottom-up signal. The studies discussed here provide evidence that predictive processing does occur and that it influences linguistic processing at different levels, including semantic and syntactic. The next crucial question for the thesis is how do prediction and disfluency overlap in language comprehension? This is discussed below when we compare and contrast predictive and attentional accounts of how disfluency interacts with language comprehension.

2.4 Disfluency

Having outlined that the focus of the current thesis is how the phenomena of disfluent speech affects the processing of language during comprehension and how this informs the debate between predictional and attentional accounts of disfluency, it is crucial to explore what these are. Therefore, the current chapter is concerned with defining the phenomena of disfluency and reviewing the existing literature, so that the reader may understand the current state of empirical research into this topic. Firstly, a discussion of what disfluency entails, comes before an overview of

the basic types of disfluent speech and the systematic differences that they display. Extra emphasis will be placed on filled pauses, the type of disfluency employed in the studies throughout the current thesis. After this, the detailing of relevant experimental findings that relate to the intersection of the current research on disfluency, prediction and attention in language comprehension, including the explicit detailing of the Predictional and Attentional accounts of disfluency processing that will be used throughout this study.

2.4.1 What are disfluencies?

A large proportion of language studies consider speakers to be consistent in their delivery, in so much as they always produce fully formed and fluent speech. Despite a high level of aptitude for speech displayed by humans, it rarely proceeds without deviation from purely fluent production; in studies of spontaneous speech it has been shown that production is affected by disfluency at a rate of 6 in every 100 words (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995). Disfluency is a term that describes a range of speech phenomena "that interrupt the flow of speech and do not add propositional content to an utterance" (Fox Tree, 1995, p. 709). Disfluency has been commonly categorised into 5 main sub-groups of different phenomena that have informed the locus of study. They have been divided up due to the differences in behaviour for both language production and comprehension that each group exhibits: repetitions, repairs, prolongations, silent and filled pauses.

Early studies focussed on the occurrence of the phenomena of disfluency, revealing that disfluency is not encountered randomly and has been shown to tend to appear before certain types of complex content, for example, open class words, such as nouns (e.g., Maclay & Osgood, 1959). Further research has shown that disfluencies also tend to occur before unpredictable lexical items, as based on contextual probability and word frequency (Beattie & Butterworth, 1979), low frequency colour naming terms (such as Pink and Orange in Dutch) (Levelt, 1983) and when naming

pictures with low-name agreement (Hartsuiker & Notebaert, 2010). Disfluency is also more frequent at major syntactic boundaries in an utterance (Goldman-Eisler, 1968).

Although disfluency has been proved useful as a group term to describe each of the above phenomena, the attributed production problems for a speaker for each type of disfluency may differ. Furthermore, these variants of disfluency have been shown to exhibit different online effects when encountered by listeners. Therefore, the variants of disfluency are discussed separately below. For the discussion of each subcategory of disfluency, it is useful to make reference to the stage of language production from which any proposed difficulty could originate. For the current thesis we will assume a three-stage model of language production (see Levelt, 1989). We do not seek to inform the debate surrounding language production and so this model is not explored in detail.

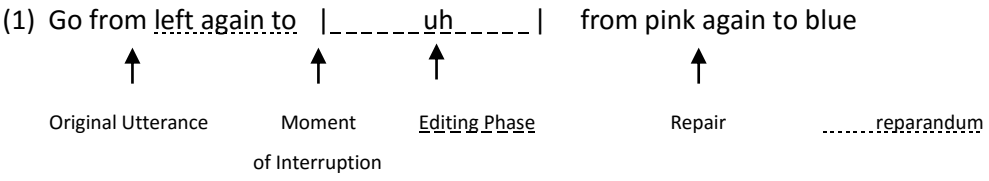
To successfully aid message transfer to a listener, a speaker must go through 3 complex stages in the production of an utterance. The first stage is the conceptualisation of content; this is the message that the speaker wishes to impart to the listener. This message draws on relevant information to guide the formulation of the concept, such as the task requirements of the upcoming utterance, the motivation for producing the utterance and knowledge stored in the long term memory. The second stage is the encoding of this message into language. This means turning the message into phonetic strings that the motor systems can execute. The final stage is articulation; this turns the phonetic plans into overt speech; at this point the speaker receives auditory feedback which is typically monitored for certain acoustic characteristics such as pitch and vowel quality.

2.4.2 Repairs

The first subcategory of disfluency discussed here is repairs; these phenomena are characterised as the break in speech when a speaker is producing an utterance but realises the preceding content to be unsatisfactory, for example, an unintended

error, and interrupts their own speech. Following this break, the speaker will try to 'repair' the previous utterance. There can be co-occurrence between repairs and hesitations, usually a silent pause upon interruption to the speech stream. However, for the current thesis we will consider this hesitation as a constituent part of the repair disfluency and separate from other hesitations, as detailed below. A repair represents the detection and subsequent fixing of an error by a speaker after the articulation stage of speech production.

Levelt's (1983) discussion of repairs provides the enduring terms used to discuss repairs. This terminology breaks the repair down into 3 notable parts. The first is the *original utterance* which begins at the end of the preceding sentence and extends to the point of the interruption in speech. The *original utterance* contains the *reparandum:* This the material which is edited during the repair. The *reparandum* can extend from a single speech sound to the entire preceding phrase. The *original utterance* boundary is marked by the *moment of interruption*. Following this moment of interruption is the *editing phase*, consisting of any delay and a possible editing term, such as a filled pause or interjection (e.g., *uh*). The final phase is the *repair* itself, this may start directly following the *reparandum* or may refer back to material contained in the *original utterance*. To put these terms to use in an example, see (1), taken from Levelt (1983) below. The *original utterance* runs from 'Go' to the *moment of interruption*, following 'to'; the *reparandum* is contained within this and covers, 'left again to' which is replaced in the repair. The *editing phase* encompasses the filled pause, 'uh'. Finally, the *repair* covers the remainder of the speech, 'from' to the sentence boundary following 'blue'.



Levelt acknowledges there to be a number of different repair types, basing the categorisation on the function of that repair, with each type representing a different

error detected by the speaker. However, there have been different taxonomies offered, with categorisation based on the structure of the repair, which do not require any pragmatic knowledge (see Finlayson, 2014).

A repair need not occur in isolation, with 17% of the recorded cases featuring 2 or more repairs within a single utterance. Hesitations, including filled pauses such as *uh* as in (1), can co-occur with repairs, commonly at the moment of interruption, allowing time for the speaker to formulate the repair (Levelt, 1983).

The repairs discussed up until this point have all featured an interruption to the speech stream, an *overt repair*. A second sub category that is much more theoretically controversial is known as a *covert repair*; this repair takes place before an utterance has begun to be articulated. These covert repairs are the result of error detection in the monitoring of inner speech that occurs between the formulation and articulation stages of production, as in Levelt's model (1989). This type of repair is controversial due to it being hard to classify using the criteria presented by Levelt (1983), as there is potential overlap with other disfluency types and planning processes. Repairs are not a focus of the current thesis and we do not further discuss the theoretical uncertainty surrounding covert repairs.

2.4.3 Repetitions

The repetition label applies to the subset of disfluency that captures the repeated use of a word, phrase or speech sound without additional content being added to an utterance, as demonstrated in (2) below.

(2) I found it...it quite hard to understand.

Not all words are equally likely to be repeated; function words, such as 'it' in the above example, are repeated more frequently than content words (Maclay & Osgood, 1959). This cannot be explained by an increased incidence of function words, compared to content words (e.g., Clark & Wasow, 1998), whose findings

showed an increased number of repetitions for function words over content words, per one thousand mentions of each.

As with repairs, there are notable distinctions between types of repetition which allow them to be classified by function. Hieke (1981) divided repetitions into *prospective* and *retrospective* categories. *Prospective repetitions* stem from anticipation of difficulty in the upcoming planning of speech. The repetitions may arise as speakers look to delay their ongoing speech to resolve the difficulty by repeating previous content, whereas *retrospective repetitions* arise as a consequence of a difficulty that has already been encountered leading to an interruption in speech. Upon the resumption of speech, the speaker repeats previous content to restore their fluency and link to the preceding utterance.

Plauché & Shriberg (1999) provided further evidence for the differentiation of repetitions into similar categories using the analysis of acoustic and prosodic information. Those repetitions labelled as prospective repetitions by Hieke are realised as *stalling repetitions* in this study, relating to when the token being repeated is prolonged. *Canonical repetitions* align with the retrospective repetitions detailed above, with the token that is to be repeated receiving prolongation. The discernible criteria that mark these repetitions as different are the divergent patterns of pause location around the repetitions, prolongation of sounds and fundamental frequency (see Plauché & Shriberg, 1999).

2.4.4 Prolongations

Prolongations are defined as speech segments whose duration extends beyond what would be considered standard in speech. Compared to other variants of disfluency, prolongations have received little empirical research. The classification of prolongations can be difficult, as judging when a segment has been lengthened beyond normal parameters can be ambiguous. From the limited study that has been undertaken, it has been demonstrated that prolongation phenomena can occur in any segment location within a word, but with a tendency towards word final

positions and more frequently in function word than content words (Eklund & Shriberg, 1998; Eklund, 2001).

A further sub-group of prolongations is the non-reduction prolongation, where a normally reduced vowel is produced, for example, "the" pronounced as *thee* rather than *thuh*. In this type of prolongation, the duration of the extended speech segment is not necessarily longer than a reduced version. Filled and silent pauses have been shown to have a greater tendency to occur following a non-reduced prolongation than following a reduced prolongation for tokens of the (*thee*/*thuh*) (Fox Tree & Clark, 1997). Bell et al. (2003) investigated the co-occurrence of disfluency and its effect on the pronunciation on an increased range of function words (*the, that, and,* and *of*). Their results showed that function words occurring in a disfluent context showed a dramatic increase in the probability of containing a non-reduced vowel sound and a notable lengthening in duration.

2.4.5 Silent Pauses

Silent pauses describe a period within an utterance when no vocalisation is produced. As with prolongations, there is room for debate within classification, as short silent pauses are a natural part of language prosody and can maintain a rhythm that aids in message transfer to a listener (see Breen, 2014), rather than a disfluency which may indicate difficulty in upcoming planning processes.

Ferreira (1993; 2007) distinguishes silent pauses into two categories based on their function in an utterance: *timing-based pauses* versus *planning-based pauses*. The former, timing based pauses, are classified as an allowance of time within a phrase, after removing the vocalisation of a word. As such, this category of silent pauses are assumed to relate to intentional prosodic breaks within speech and will not be considered as disfluency for the current thesis. In contrast planning based pauses are unintentional interruptions to spontaneous speech that arise as a consequence of production difficulty with upcoming speech, which meet the definition of a disfluency.

2.4.6 Filled Pauses

The final subset of disfluency is filled pauses (fillers). This variant of disfluency is of greater importance to the current study due to it being the phenomenon under investigation in the empirical research contained in this thesis. A filled pause is defined as a gap in an utterance that is filled with vocalisation, most notably but not limited to *uh* and *um*; in British English, *uh* is written as *er*.

Filled pauses have been the focus of much psycholinguistic work. It has been suggested they are categorised as discrete from each other, with each signalling a different intentional message to the listener and with equivalent pairs of filled pause phenomena found across languages (see Clark & Fox Tree, 2002). In Clark and Fox Tree's (2002) view, filled pauses are considered as similar to conventional words in English, representing a signal to the listener that a delay is upcoming. The distinction between *uh* and *um* is based on the duration of interruption that the speaker wishes to convey to the listener, either a minor or major predicted delay, respectively. This claim was based on them finding that in a spoken corpus, a longer silent pause followed an *um*, than an *uh*. They also make claim that each of the filled pauses can occur with a prolongation, a lengthening of the schwa vowel, that signal to the listener that the difficulty is ongoing. The pattern of difference between *uh* and *um* seen for this study has been demonstrated in other empirical research (see Fox Tree, 2001; Barr, 2001).

The Clark and Fox Tree account is not uncontentious, with criticism of the method that they used to measure the filled and following silent pauses (see Schnadt 2009; Corley & Stewart, 2008; O'Connell & Kowal, 2005); the units of prosodic stress employed were conventions coded into the corpus by transcribers and lacked objectivity. However, as speakers naturally differ in speech rate, this variance in speech rate could affect the length of duration, for both the proposed minor and major delays. Therefore, the advantage of this subjective measure of filled and silent pause length is its sensitivity to the individual speaker's natural speech rate.

O'Connell and Kowal (2005) used a corpus of six media interviews with Hilary Clinton to accurately measure the duration of *uh*, *um* and the co-occurring silent pauses present. Their results provided striking differences to those found in Clark and Fox Tree: they found limited occurrence of any duration of silent pause following either type of filled pause and there was no significant variation between the duration of silent pause following either filled pause variant. They also challenged the suggestion that filled pauses act as signals to the listener. As Hilary Clinton is deemed to be an expert public speaker, they propose that she should be able to exert control over the production of filled pauses to effectively signal her intentions during speech. However, if disfluencies are viewed as a negative phenomenon in public speaking, it is likely that Hilary Clinton, as a highly trained professional speaker, would seek to avoid any form of disfluent pause. Furthermore, if filled pauses are an audience design feature, you might expect that the presence of an interlocutor would increase the rate of filled pauses, as a speaker uses disfluency to signal to a listener about any upcoming difficulty. However, in a separate study in which the frequency of filled pauses were measured there were no differences seen in the distribution of disfluency between speech from monologues and dialogues (Finlayson & Corley, 2012).

In the face of the current research, it remains unclear as to whether of *uh* and *um* are intentionally employed to indicate different lengths of upcoming delay. However, the evidence does point to both *uh* and *um* occurring with a delay of different duration. The view taken for the rest of the thesis will be that *uh* and *um* represent subcategories of the filled pause phenomenon.

2.5 Disfluency and Comprehension

Having examined the variants that make up disfluency, this chapter now investigates the effect of disfluency on the language comprehension processes of a listener. Presented first is background on general disfluency processing during

comprehension. After this, we focus on filled pauses and how they have been shown to affect the listener.

How do disfluencies, interruptions to a fluent speech stream, impact upon online language comprehension? One option is that they represent non-linguistic noise, which a listener needs to remove to understand the message transferred in the linguistic content and let comprehension processes proceed, as in fluent speech. There has been some support for this account previously; Martin and Strange (1968) stated that encountering disfluency in speech disrupts the processes underlying speech perception and it is, therefore, filtered out. In Levelt (1989) disfluent speech presents a continuation problem, meaning that listeners must edit out the disfluencies to successfully comprehend speech. Furthermore, as listeners are poor at accurately locating disfluencies in a sentence that they just heard (Lickley & Bard, 1996), this could suggest that listeners are not comprehending disfluencies in the same manner as other linguistic material.

However, this standpoint has since been seriously challenged by a second account that believes disfluency does not always hinder comprehension and can be beneficial in aiding a listener (e.g,. Brennan & Schober, 2001; Fox Tree, 2001). Brennan and Schober (2001) found that certain disfluencies helped compensated for mishaps in speech. At the least, disfluency has been shown to have clear effects on comprehension, influencing the expectations of upcoming content (Arnold et al., 2007, 2004), parsing of garden-path sentences (Bailey & Ferreira, 2003), attenuation of context dependent word integration (Corley, MacGregor, & Donaldson, 2007) and speeding up of word recognition (Corley & Hartsuiker, 2011).

Also, it is worth noting the impact disfluency has on the judgement of attributes of a speaker, for example, a tendency to believe that general knowledge answers that follow disfluency have lower confidence ratings (Brennan & Williams, 1995) and are more likely to form worse impressions of speakers producing *um* (Christenfeld, 1995). These studies demonstrate the attribution of certain properties to a speaker

based on the fluency of their performance. Overall, it is clear that disfluency can affect the comprehension processes in a number of ways.

2.5.1 What do filled pauses mean for comprehension?

Here we discuss in more detail how filled pauses have been shown to affect language comprehension, in both the short and long term. As attested to above, a clear effect in the literature is that upon encountering the filled pause, *uh*, listeners show a bias for unknown or discourse new referents (e.g., Arnold et al., 2007, 2004). These visual world studies are presented in more detail below. Bosker, Quené, Sanders, and de Jong (2014) showed that similar results held following an *um*, with listeners demonstrating a preference towards low-frequency referents over high-frequency objects. Additionally, there has been evidence that listeners are flexible in the attribution of disfluency effects based on their knowledge of speakers: when hearing instructions from multiple speakers, they show sensitivity to tracking whether objects are discourse new or not for each speaker, showing a tendency towards the reference of unmentioned objects for a speaker who produces a disfluent utterance (Barr & Seyfeddinipur, 2010). Similarly, when listeners are told that a speaker had a condition that led to difficulty in naming objects, they did not demonstrate a tendency towards referents without a conventional name as is seen for typical speakers (Arnold et al., 2007). This suggests that listeners infer the cognitive state of the speaker in relation to this situation that is speaker specific and this attenuates the impact of disfluency on their predictions.

Filled pauses have also been shown to affect sensitivity to prediction in ERP studies; Corley et al. (2007) investigated how filled pauses affected the semantic integration of either predictable or unpredictable target words into the preceding linguistic context. An example context, "Everyone has bad habits, mine is biting my tongue", could be shown with either the unpredictable target "tongue", or alternatively with the predictable ending "nails". In the disfluent conditions, the filled pause, "er" was located immediately before the target word. As well as finding the established effect

of increased amplitude of the N400 component following unpredictable words in the fluent condition, they also demonstrated that following *er*, listeners' N400 responses to contextually unpredictable words were significantly reduced. This suggests that the disfluency is impacting the ease with which an unpredictable word is processed; disfluency appeared to have modulated the listeners' predictions about the upcoming content. Additionally, participants were then asked to take part in a surprise memory test, featuring visually presented words that had may have appeared in the previous listening task. They were tasked with selecting whether they had seen those words before. Those which had co-occurred with disfluency at the previous stage were recognised more often by participants, showing a memory facilitation effect. This effect takes place over a relatively longer term, compared to the original listening task, suggesting that disfluency can have lasting impacts on the understanding of speech. Taken together, these studies provide evidence that both types of filled pause used in the current thesis can impact on the expectations of upcoming content.

Fox Tree (2001) investigated the influence of both *uh* and *um* on word recognition in Dutch and English across separate experiments. Participants were tasked with monitoring recorded spontaneous speech and pressing a button upon hearing a target word. The reaction times of the participants were measured. In all trials, the target words were recorded being preceded by filled pauses, half *uh* and half *um*. However, following this 50% of all trials then had the filled pause removed, creating a fluent condition. The results demonstrated that when a target word was preceded by *uh*, participants were quicker to detect the target, compared to the fluent condition. In contrast when a target word was preceded by *um* there was no divergence between the reaction times to detect a target word compared to a lack of filled pause. These findings suggest that *uh* and *um* are causing listeners to respond differently, possibly as a consequence of each disfluency representing a different behaviour in relation to the duration of planning difficulty, rather than just wholly being based on the different time course of each variant. These results have not gone

unchallenged, whilst Corley and Hartsuiker (2011) found that *um* also facilitated identification of a target word, this effect was also seen for an equivalent silent pause, suggesting that it is simply the increased delay that is responsible for the heightened ability for word identification. The resultant *temporal delay hypothesis* is discussed below.

Overall, filled pauses have been evidenced to provide a range of comprehension effects. The majority of studies provide support for the viewpoint that when heard, listeners infer disfluency to mean that the speaker is facing difficulty. However, this is contentious as some evidence has shown that filled pauses produce the same effects to those seen for delays which feature no vocalisation, silent pauses. Below 3 current accounts of disfluency processing are explored.

2.5.2 Disfluency and Comprehension: What processing underlies these effects?

So far we have simply stated how disfluency has been shown to impact upon the comprehension processes we have not investigated the mechanisms underlying these effects. Below we highlight the theories and research that point to a certain account of disfluency processing, primarily following hesitation phenomena. The current thesis aims to try and differentiate between predictional and attentional accounts from the impact of disfluency, reconciling the effects seen in the literature with one account and providing support with further empirical study. These two accounts are explored and we define what they mean for the current thesis.

2.5.3 Disfluency and Prediction: Predictional Account

As evidenced below in the visual world section, disfluency has been shown to have a notable impact on predictive processes in language comprehension (e.g., Arnold et al., 2007; Heller, Arnold, Klein, & Tanenhaus, 2014). So how does disfluency interact with underlying expectations to create the effects detailed above? Firstly, we have outlined what we define as predictive processing during language comprehension above. A predictional standpoint for disfluency processing (e.g., Arnold et al., 2007;

2004; Heller et al., 2014) suggests that upon encountering disfluency, a listener infers the speaker to be experiencing difficulty. This is a difficulty similar to that seen in situations where speakers tend to become disfluent, such as when they are experiencing cognitive load (Bortfeld et al., 2001; Brennan & Schober, 2001) or in the face of increased difficulty in lexical retrieval, for example, when trying to produce a word that is contextually unpredictable or low frequency words (Beattie & Butterworth, 1979). Listeners build up patterns of disfluency distribution information that inform their expectations of upcoming content. As noted above, the reliance on this distributional information is flexible and can be modulated by other knowledge that influences the cognitive representation of the speaker, whether this is that they have difficulty naming objects (Arnold et al., 2007), there are multiple speakers each with a different set of discourse new and old objects (Barr & Seyfeddinipur, 2010) or that the speaker is non-native and may have a variable pattern of disfluency (Bosker et al., 2014).

Heller et al. (2014), detailed in the discussion of the visual world paradigm below, demonstrated the flexibility of the online disfluency processing mechanism in response to situational and speaker specific contextual information. They showed that instead of listeners directly associating disfluency with properties of objects, for example, a lack of conventional name, they used situation-specific inferences to guide their predictions for upcoming referents. However, they also revealed limitations to these inferences, as listeners showed a lack of sensitivity to assumed speaker knowledge of referent names when it diverged from their own experience of the names. This contrasts with the results seen in Barr and Seyfeddinipur (2010), who showed that following a disfluency, listeners were sensitive to speaker specific knowledge of discourse mentions for referents. In summary, the predictional account of disfluency processing relies on a probabilistic attribution of speaker difficulty, coupled with situational and speaker specific knowledge to infer the cognitive state of the speaker and uses this information to update expectancies for upcoming content.

2.5.4 Disfluency and Attention: Attentional Account

An alternative account of disfluency processing during comprehension has been proposed by Fox Tree (2001). Revisiting the study, detailed above, the results showed that following *uh*, participants were quicker to identify a target word than in the related condition that featured a silent pause of the same duration. Fox Tree interprets this finding as support for an attentional account, in which this filled pause "heightens listeners' attention to upcoming speech," (p. 325). However, *um* revealed a lack of significant divergence in participants' reaction times compared to the silent pause condition. Fox-Tree proposed the reason for this lack of divergence as being that *um* is thought to represent a longer upcoming delay in speech. Fox Tree further suggests that orienting attention would be impractical when the time course of the resumption of speech is unknown. However, Corley and Stewart (2008) proposed a different explanation, namely that the duration of the silent pause from the removed *um* represents a delay that extends beyond a normal gap in fluent speech and is notably longer than for either the filled or silent variants of *uh*. Therefore, the silent pause in the *um* condition could have been comprehended as disfluent or processed in a manner divergent from typical fluent speech.

Further support for an attentional mechanism in disfluency processing was demonstrated by Collard, Corley, MacGregor, & Donaldson's (2008) ERP experiment. In this study, participants heard recorded speech which featured intermittent single words which had been acoustically manipulated to alter a characteristic of that word, for example amplitude, so that it audibly differed from the surrounding speech. Half of the utterances included an additional filled pause, *er*, preceding the manipulated target word, the remaining half were fluent. In the fluent condition, the manipulated target words produced predictable MMN (mismatch negativities) and P300 components, when compared to unaltered utterances, whereas, following the disfluent utterances, the MMN component was still seen but the P300 was greatly reduced. The P300 component is associated with the orienting of attention to novel stimuli. Therefore, this reduction in the P300

following disfluency suggests that participants were already attending to the incoming speech. These results support the viewpoint that following a disfluency, listeners are orienting their attention to the upcoming content and this explains previous facilitation effects seen following filled pauses (e.g., Brennan & Schober, 2001; Fox Tree, 2001). A possible reason behind this facilitation is that the disfluency causes listeners to rely on the incoming speech signal, bottom-up information, to resolve the comprehension difficulty posed by the interruption to the speech, whilst the increased attentional resources allow a quicker recognition of following linguistic content.

The attentional account proposes that following a disfluency predictional processes are reset and the additional attentional resources are used to focus on bottom-up processing to facilitate comprehension. This account contrasts with the predictional viewpoint, which attributes the effects seen post disfluency as consequence of probabilistic speaker inference, top-down contextual information, to alter expectancy about upcoming content. The attentional mechanism need not be necessarily mutually exclusive from the inference based, predictional account of disfluency processing (e.g., Arnold et al., 2007).

2.5.5 Temporal Delay Hypothesis

Although not explicitly tested in the current thesis, a third disfluency processing account has been proposed, the temporal delay hypothesis (e.g., Corley & Hartsuiker, 2011). Exploring this account could improve our understanding of the other accounts investigated in the current thesis and add to our knowledge about the parameters that can affect disfluency processing. Across three experiments, Corley and Hartsuiker (2011) tasked participants with viewing a pair of images whilst listening to instructions that asked them to select one with a button press, for example, "now please press the button for [*um*] book, please". In each experiment, they tested the influence of a different delay marker preceding the target object versus a condition where a delay of equal length appeared earlier in the utterance. The delays were of equivalent durations but marked with different phenomena

between experiments: the first was the filled pause, *um*; the second was a silent pause and the third employed an artificial tone. Their results showed a facilitation effect in word recognition following all delays located immediately pre-target compared with the control condition, where the delay had occurred earlier in the utterance. This account suggests that there should be no variation in comprehension effects between silent and filled pauses. There has been support provided for a temporal delay account in other studies: Bailey and Ferreira (2003) showed equivalent disambiguation of garden path sentences for both the filled pause *uh* and background noise for a comparable duration. Similarly, Corley and colleagues showed an attenuation effect on the N400 component for contextually unpredictable words following both filled pauses (Corley et al., 2007) and silent pauses (MacGregor, Corley, & Donaldson, 2010). However, this account also contrasts with a number of findings which show divergence in effects following a filled pause compared to a silent pause: In Fox Tree (2001) the findings suggest that the vocalisation during a filled pause is driving the observed effects due to the facilitation effects of *uh* over a silent pause of equivalent duration. Additionally, Barr and Seyfeddinipur (2010) found a difference in the size of effect following an *um* compared to a condition where it was replaced with noise.

Two reasons, proposed in Corley and Hartsuiker (2011), as to why a delay could be responsible for the facilitation effects demonstrated, both stem from the extra time allowance given to listeners to process during comprehension. The first is the additional time allows listeners to implement top-down processing and use it to aid visual recognition in the search for a referent. This links the temporal delay hypothesis to results that support a predictional account of disfluency processing (e.g., Arnold et al., 2007; Bosker et al., 2014; Heller et al., 2014). In these studies, the delay of a filled pause could allow extra time for linguistic top-down expectancy and inferential speaker processing pertaining to the upcoming referent in the scene to reach the visual search behaviour resulting in the effects seen. However, the predictional account suggests that listeners are basing their expectations upon

distributional information of disfluency occurrence, with a certain type of planning difficulty, notably filled pauses. This cannot be reconciled with the Corley and Hartsuiker (2011) findings, as the facilitation effect demonstrated does not necessitate the need for a filled pause, merely a delay. The second reason is that the increased time allows listeners to orient their attention to upcoming speech, facilitating the recognition of linguistic content following on from the delay. This ties in with an attentional account of disfluency processing, as evidenced in Collard et al. (2008) who showed that the delay associated with a filled pause resulted in a reduced amplitude of mismatch negativity and P300 component during ERP measurement.

In summary, the temporal delay hypothesis can readily overlap with an attentional account and support predictive effects seen for disfluency processing but it is much harder to reconcile with a predictional account. However, the evidence on whether the facilitation effects following a filled pause hinge on the audible production of a disfluency that leads to an inference of difficulty from a speaker or a delay that can remain silent is contentious and far from certain. The temporal delay hypothesis does not form a central tenet in the current thesis but at least it provides evidence of another useful property of disfluency that can inform our understanding of how disfluency impacts upon language comprehension.

2.6 Visual Word Studies

The visual world-paradigm, as typified by Altmann and Kamide (1999), has proved a rewarding method for exploring language processing (see Altmann & Kamide, 2007). This paradigm is used in the current thesis to further investigate the influence of disfluency on language comprehension and to illuminate the debate between prediction based and attentional accounts. Below we highlight the validity of visual world studies to inform about the online processing of language and how this methodology has provided insight and evidence for prediction in the processing of language. Also explored is the current empirical evidence relating to the effects

noted upon encountering disfluency in the speech input. Other studies have examined the intersection of disfluency and prediction in a visual world setting and these are discussed in relation to the aims of the current thesis.

2.6.1 What is a visual world study?

A typical visual world experiment presents a participant with a visual scene consisting of a number of pictured objects, whilst being simultaneously presented with spoken language. Taking an example sentence used in Altmann and Kamide (1999), 'The boy will pick up the ornate red vase'. A listener will be constrained by the linguistic entities heard if processing proceeds incrementally. The noun phrases, 'The boy' and 'the ornate red vase', limit the possible referents in the real world. The verb narrows down to entities that can be picked up. Each adjective constrains potential referents in the real world to those that meet the updated criteria, namely, those objects are first ornate and then red. This precedes the onset of the final noun, 'vase', so if listeners are processing in an incremental manner, after hearing the ongoing speech stream, they should only look towards the pictures that are still valid from the constraining linguistic context.

2.6.2 Are listeners' eye-movements sensitive to linguistic context?

The visual world paradigm has provided empirical evidence of incremental word-by-word sensitivity to linguistic content. Eberhard, Spivey-Knowlton, Sedivy and Tanenhaus (1995) first developed the visual world methodology that is known as such today. Their primary concern when creating this paradigm was to examine the temporal sensitivity of the processing that underlies unfolding language. This built on earlier work that had shown that when participants viewed a visual scene whilst simultaneously being presented with speech, participants eye-movements were sensitive to the linguistic content conveyed within the spoken language (Cooper, 1974).

Additionally, in other experiments participants have to attend to linguistic input in a way that they would not outside of the lab. A positive taken from the visual world study is that it allows language to be understood in a more natural way, whilst having a measure, eye-movement that is sensitive to the comprehension processes, rather than using a secondary task, such as cross-modal priming.

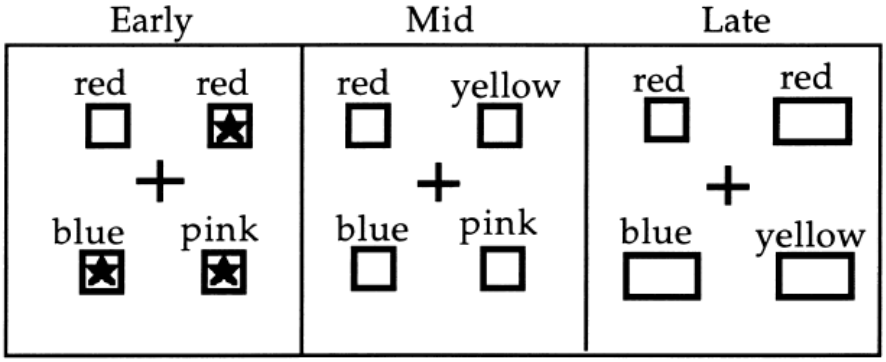


Figure 2.2: Example displays (taken from Sedivy et al. 1999) that show the contrasting manipulation points for a spoken instruction, '*Touch the plain red square*', to be resolved. In each condition, the point at which the instruction allows a unique referent varies.

They examined whether participants demonstrated incremental processing in referring expressions, by asking participants to manipulate objects presented in a visual scene. Participants heard an instruction, such as, 'Touch the plain red square'. They created early, middle and late points of ambiguity resolution for the instructions by manipulating the colour and presence of a star on the objects in the visual array, as in the early condition, participants could disambiguate the object from the adjective 'plain', as in this visual scene there were no other plain shapes. The middle condition could be disambiguated at 'red', as all shapes were plain but there was only a single red object in the scene. In the late condition, participants could only disambiguate the shape upon 'square', as all shapes were plain, there were 2 red shapes but only a single red square. Their findings showed that participants were able to extract the relevant online information from the speech

stream resulting in looks to the target referent, time locked to the point of ambiguity resolution.

Sedivy, Tanenhaus, Chambers and Carlson (1999) provided further evidence for online sensitivity to linguistic content in visual world experiments. They extended the paradigm to look at the context-dependent effect of semantic interpretations of pre-nominal adjectives. For example, for participants responding to 'tall glass' they manipulated the typicality of the picture of the glass, whether the picture was a good or poor example of a tall glass. They controlled for context, by either having only a single glass that matched the referring expression heard or a second competitor glass; this competitor glass contrasted with the property of the adjective (e.g., a smaller glass). Meaning that in the context of the visual scene, the 'tall glass' picture could be disambiguated before the noun onset. Their results again showed incremental processing of the adjective that was sensitive to the context-specific contrast in property of the picture before the noun onset. Even unfolding language at word level is enough to influence participants eye-movement behaviour (Allopenna, Magnuson, & Tanenhaus, 1998), participants can begin to evaluate a referring expression, such as, 'the table' within 200ms of noun onset. However, the visual world paradigm is not without its limitations. It is clear that in arrays such as those detailed above the number of objects and the situation presented does not reflect the variability in visual scenes outside of a lab situation (as noted in Allopenna et al., 1998) or mimic the real world experience of using language in an ecologically valid way in the real world. Taken together, these studies prove without contention that the visual world paradigm, as a tool, is clearly useful in tracking listeners' sensitivity to online language processing.

2.6.3 Visual World and Prediction

In the current thesis we aim to explore the effect of disfluency on language comprehension, to weigh in on the debate between prediction based and attentional accounts. The visual-world paradigm has also proved useful in exploring predictive processing. To be clear about the distinction being made here, the prediction we

discuss here is the matching of incoming speech or language against expectancies derived from top-down information, rather than comprehension solely being built from the bottom-up information received alone. In a highly influential study, Altmann and Kamide (1999) employed a visual world paradigm to demonstrate semantic prediction encountered at the verb in language comprehension. An example scene (Figure 3) is comprised of a boy, a cake and two distractor items. Participants hear two variants of a sentence context, either 'the boy will eat the cake' or 'the boy will move the cake'. At the onset of the verb in the 'eat' condition, participants can predict the upcoming target theme that will subsequently be heard, using the selectional restrictions of the verb to select the cake, as it is the only edible referent in the scene. Whereas, in the move condition, due to the ambiguity of other moveable referents being available in the scene, the point at which the target referent, cake, can be uniquely identified occurs after the onset of the noun. Their results demonstrated this, with looks to the target occurring before the onset of 'cake' in the 'eat' condition and after the onset in the 'move' condition. This provides evidence for predictive processing as the fixations occurred before the onset of the target noun. These results support the use of contextual information to create expectancy between the verb, e.g., 'eat' and the semantic features of a referent theme, namely whether objects in the scene are edible.

Further study by Kamide, Altmann and Haywood (2003) set out to show that the effect seen in Altmann and Kamide (1999) was not driven by the associative relationship between the verb and the subsequent speech alone. Using a scene that depicted a man, a young girl, a motorbike and a carousel whilst participants heard either 'the man will ride' or 'the girl will ride' they demonstrated that the predictive processing was sensitive to the combination of the subject of the phrase and the verb. Following, the 'man' condition participants showed an increased proportion of looks towards the motorbike, compared to the carousel. Whereas, after hearing, the 'woman' sentence participants showed the reverse, an increased proportion of looks towards the carousel over the motorbike.

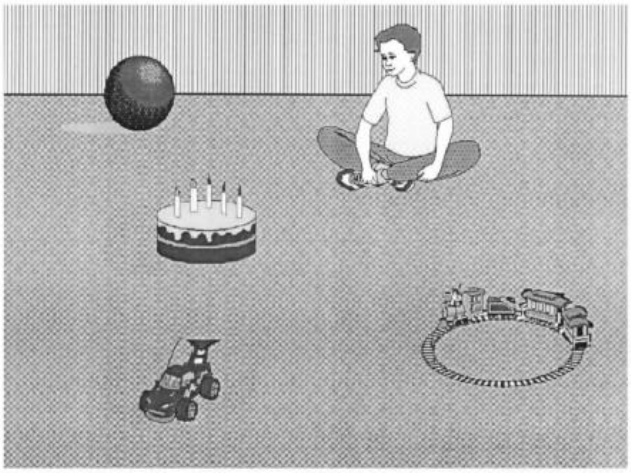


Figure 2.3: An example scene from Almann and Kamide (1999). Participants heard either: *'the boy will eat the cake'* or *'the boy will move the cake'*.

This shows that participants were still predicting based on the selectional restriction of the verb, something to ride, but they were additionally influenced by the semantic context of the preceding subject. This provides evidence that the predictive effects seen cannot be explained by the association of a semantic field between a verb and a referent. Additionally, Altmann and Kamide (2007) demonstrated that participants were sensitive to the selectional restrictions created by the tense of a verb. Recent work has sought to find the limitations of predictive processing using the visual world paradigm as a measure of the underlying processing taking place (e.g., Kwon & Sturt, 2014). Taken together, the research reviewed here provides clear evidence for predictive processing during language comprehension, also, showing that the visual world paradigm is adept at capturing these expectancy effects.

Participants are also sensitive to other forms of top-down information that do not originate from the content heard, such as the discourse context, notably, whether an item is new to the discourse or has featured previously (Kaiser & Trueswell, 2004). Another demonstration of a prediction effect from non-content contextual information stems from the prosodic information of the speech stream (Weber, Grice, & Crocker, 2006): the prosodic cues during speech gave rise to expectancy

about the grammatical function of an upcoming referent. Participants have also been shown to be sensitive to predicting upcoming events (e.g., Knoeferle, Crocker, Scheepers, & Pickering, 2005).

2.6.4 Visual World and Disfluency

In the current thesis, we are concerned with the comprehension processes when disfluency is encountered in the speech stimuli. Disfluency represents the addition of extra information added to the speech stream, as discussed above. The visual world has also been proved to be effective as a method of exploring the linguistic processing that occurs during language comprehension with this speech input. Arnold, Tanenhaus, Altmann and Fagnano (2004) incorporated a filled pause, 'uh', into a visual world paradigm to investigate whether the inclusion of disfluency changed participants' eye-movement behaviour between a referent previously mentioned in the discourse or discourse new referents. A typical scene consisted of two objects that were cohort competitors, meaning that they shared the same initial sounds, for example, 'candle' and 'camel' and two distractor pictures that shared no phonetic overlap to two target competitors (Figure 4). Participants viewed a visual scene whilst hearing instructions with the goal of moving one of the objects around the scene. The instructions always consisted of two sentences: the first referred to one of the competitor objects, marking it as discourse established and the remaining competitor object as discourse new, for example, 'Put the camel below the grapes'; The second sentence then asked the participants to manipulate a competitors objects but could be realised in a fluent or disfluent condition, for example, *'Now put [the/ thee uh] candle..*'. The target of the second sentence could either be the discourse established or discourse new object. This created a 2X2 design with fluency (fluent/disfluent presentation) and discourse status (established/new). The results from this study revealed that following the presence of a disfluency participants made more fixations on the discourse new object, compared to the fluent presentation. This finding showed that following a disfluency participants adjusted their expectancy about the upcoming object. That is to say that their online

comprehension processes proceeded in a different manner following disfluency.

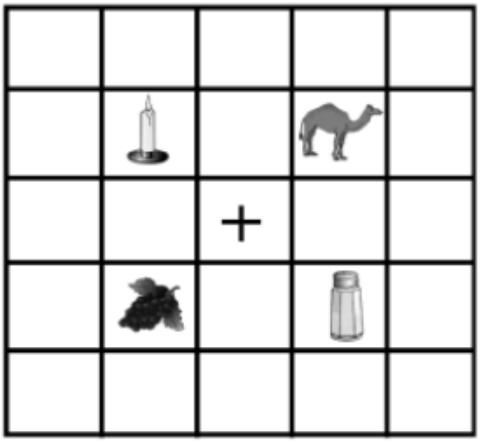


Figure 2.4: A visual array taken from Arnold et al. (2004).

In Arnold, Kam and Tanenhaus (2007) there was a replication of disfluent instructions causing a change in expectancy for an upcoming referent. In this study, Arnold and colleagues used visual scenes containing 2 familiar and 2 unfamiliar objects. The unfamiliar objects were harder to describe due to a lack of conventional name. The disfluency employed in this study matched that used in the earlier study, the filled pause, *uh*. In the face of disfluent instructions participants fixated more on the unfamiliar objects. This result provided another instance of on-line comprehension processing that has demonstrated a change in expectancy for an upcoming object upon encountering disfluency in the speech steam. Arnold and colleagues suggest that this effect may be facilitated by either participants sensitivity to speaker based challenges in production such as, describing something difficult, planning a new utterance or distraction. This account forms the basis of the prediction based account that we test in the current thesis. This account and its theoretical groundings are discussed in more detail above.

Corley (2010) investigated the influence of speech repairs, another form of disfluency, on language comprehension. Building on the paradigm used in Altmann & Kamide (1999), it set out to test whether the predictive processing explicitly took

into account the repair. Corley created 4 conditions, including the 2 fluent conditions (i & ii) that featured in the original Altmann & Kamide paper:

i) The boy will eat the cake (restrictive verb)
ii) The boy will move the cake (nonrestrictive verb)
iii) The boy will eat and move the cake (conjunct)
iv) The boy will eat- *uh*, move the cake (repair)

Using these conditions the fixation behaviour on the determiner could distinguish whether participants were updating their predictions following a repair. If the prediction processes do not explicitly take the repair into account then condition iii and iv should provide a similar pattern of fixation results to those seen in condition i, as the restrictive verb will still influence predictions and narrow down to target themes that have to be edible. In contrast, if the repair is monitored explicitly then there should be divergence between condition iii and iv as participants update their predictions. This means that for condition iii, it would be restrictive (as in condition i), as the selectional restrictions of each verb are joined together, so the target must still be something edible to meet the restriction of the 'eat' verb. However, if in condition iv participants are updating their expectancies upon encountering the repair, namely, overriding the initial verb that was heard; then following this disfluency their pattern of fixation behaviour should follow condition ii as 'move' is nonrestrictive. The results showed that listeners were explicitly attending to the repair, causing them to update their predictions. In the conjunct condition (ii) participants made fixations that followed the pattern seen for restrictive verbs (i). Following a repair (iv), participants' fixations patterned with the nonrestrictive verb (ii). This lends support to predictions during comprehension processing being updated incrementally using the cues from the available linguistic and non-linguistic context. This study is interesting because if Arnold and colleagues' account of listener sensitivity to speaker based challenges is correct, then the speech repair should clearly signpost a production difficulty leading them to update their inferences following the incidence of the repair, which is what is found in the

results. The findings here also show that participants are sensitive to speech repairs, another form of disfluency, during the online processing of language during comprehension and that the previous results with *uh* are not specific to this form of filled pause.

Recently, Heller, Arnold, Klein, and Tanenhaus (2014) examined whether upon encountering disfluency, the inferential processing undertaken by listeners is regulated by objects that have set properties, notably, not being mentioned previously, lack of a conventional name or requiring a longer description. Alternatively, following a disfluency listeners could adapt situation-specific inferences that are not tied to object properties associated with disfluency. Using an artificial mini-lexicon, that was learned by participants, they aimed to dissociate disfluency from the referent properties it often co-occurs with by having the objects with learned names contrast with objects that are new to the discourse. They created scenes that featured pairs of named objects (names taken from the learned lexicon) and unnamed objects. Each member of the pairs looked the same as the other object in that pair, except one of each pair of objects was the same colour (e.g., red) and the remaining object in each set another (e.g., blue). They hypothesised that the newly learned named objects would be perceived as difficult to produce, based on previous research that newly learned words take longer to produce than known words (e.g., Costa, Santesteban, & Ivanova, 2006). The unnamed pairs met the criteria of associative properties with disfluency, namely that these objects had not been named before, requiring a longer description. Participants then heard either fluent or disfluent instructions to manipulate one of the objects in the scene, for example, '*Click on the/thee uh red..*'. Following a disfluent instruction, if the participants are using situation-specific adaptation for their inferential processing then it would be predicted that they would fixate more on the named objects, as they may expect the hard to produce newly learned names following a disfluency. However, if the inferences made following disfluency are tied just to object properties then participants should update their expectancies to the unnamed objects. Their results provided evidence for listeners being able to spontaneously

update their inferential processes to adapt to situation-specific inferences. However, this study also showed the limitations of the online flexibility of these inferences, as the listeners were not sensitive to the speaker's assumed knowledge of the objects learned names when they diverged from their own. Despite this lack of complete situational awareness, the visual world study shows the flexibility of online disfluency processing and that eye-movement behaviour can differentiate between fine-grained distinctions.

In order to answer the question of whether a prediction based or attentional account can better explain the disfluency effects seen in language comprehension we first needed to find suitable methods to empirically test between the fine-grained differences in empirical predictions stated by each account. Based on the evidence of the research above, a visual world paradigm has clearly been shown to be sensitive to capturing the effects of a variety of linguistic and non-linguistic factors, including predictive processing and the processing linked to disfluency in the speech stream. This validates the paradigm as a suitable method to differentiate between the accounts of disfluent language comprehension.

2.7 Conclusion

Disfluency is a phenomena that is relatively frequent in everyday speech and has been the focus of much empirical research. The current chapter provides an overview of the subgroups of disfluency: repairs, prolongations, repetitions, filled and silent pauses. We had special interest in filled pauses, showing that there is ongoing debate surrounding whether the production of *uh* and *um* signal delays of varying length to the listener (Clark & Fox Tree, 2002; Fox Tree, 2001). We take the view that each filled pause represents a separate sub-group.

We documented the numerous effects of disfluency during language comprehension, with increased focus on the filled pause variants of disfluency that are used in the following research and especially how these pertain to the 3 accounts of disfluency processing we explored: The predictional (e.g., Arnold et al., 2007),

attentional (e.g., Collard et al., 2008) and temporal delay (Corley & Hartsuiker, 2011) accounts.

We have provided clear evidence that listeners' attention can modulate the use of top-down and bottom-up processing depending on the situation or task demands; both attending to the fine grained acoustic input (e.g., Mirman et al., 2008; Pitt & Szostak, 2012) or increased use of contextual knowledge when under cognitive load (e.g., Mattys & Wiget, 2011). The relevancy of the discussion of attentional effects and how this informs models of speech perception has also been explored in the current chapter. Additionally, we have provided enough evidence to take the stance that prediction and expectancy are an implicit part of language processing (e.g., Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Federmeier et al., 2007). We have also spoken to how the visual world paradigm is an effective measure of linguistic sensitivity for both prediction (e.g., Altmann & Kamide, 2007) and disfluency (e.g., Arnold et al., 2007; Heller et al., 2014). In summary, we have explored the central themes of the thesis: disfluency, attention and prediction. Next, we use this knowledge to empirically test our research aim of testing between the predictional and attentional accounts of disfluency processing during comprehension.

CHAPTER 3

Experiment 1

3.1 Chapter Overview

In the previous chapter, we have shown the validity of the visual world methodology to capture sensitivity to linguistic manipulations (e.g., Altmann & Kamide, 1999). Disfluency processing during language comprehension is the central theme of the current thesis and we have outlined and detailed both the Predictional (e.g., Arnold, Kam, & Tanenhaus, 2007) and Attentional (e.g., Collard, Corley, MacGregor, & Donaldson, 2008) accounts for the effects that have been evidenced occurring with disfluency. In the current chapter, we present the first experiment, in which we use the visual world paradigm to explore language processing following a filled pause, with a view to differentiating between the Predictional or Attentional accounts.

3.2 Introduction

It has been shown that there are a number of complex predictive processes that take place during language comprehension (see Federmeier, 2007; Pickering & Garrod, 2007). Listeners have demonstrated sensitivity to the semantic information contained in unfolding linguistic input. In reading, the semantic context of a preceding sentence exerts influence on a sentence final target word: Participants are quicker to recognise letter-strings as words from predictable context (e.g., Schwanenflugel & Shoben, 1985).

The visual-world paradigm has proven its usefulness in measuring participants' online predictive processes in response to spoken stimuli by tracking eye-movement behaviour. For spoken language, there is clear evidence that language users are influenced by online contextual information in an incremental manner as measured

by their eye-movement behaviour, with the time-course of eye-movement following the pattern of linguistic input (e.g., Altmann & Kamide, 2007; Kamide, Altmann, & Haywood, 2003). Participants will build expectancies for upcoming content, looking towards the pictorial representation of likely referents in a visual scene while hearing sentential stimuli (e.g., Altmann & Kamide, 1999). The expectancies created by the content of what is heard are not driven by straightforward associations between words encountered in the previous context and the upcoming words (Kamide et al., 2003). Support for prediction is also provided from evidence in ERP studies. Kutas and Hillyard (1984) found that when a semantically unexpected noun, such as 'coffee' was heard following a biasing sentence context, 'He liked lemon and sugar in his..', it resulted in an increased N400 effect compared to a highly predictable noun, such as 'tea'. This N400 effect has been replicated whilst controlling for different patterns of variation in the sentential context and target word. For example, it has been shown that there is a reduced N400 effect for anomalous words semantically related to a predicted word for the preceding context (Federmeier & Kutas, 1999). Participants predict the form of upcoming words, with listeners demonstrating sensitivity to a mismatch effect at the point where their expectation of a particular phoneme and the realisation of a different phoneme diverge (DeLong et al., 2005).

Prediction does not apply exclusively in a semantic domain; the syntactic information included in a context can also drive expectancy for upcoming content (Lau et al., 2006; Van Berkum et al., 2005; Yoshida et al., 2013). Van Berkum et al. (2005) found that Dutch listeners were sensitive to a grammatical gender mismatch between an adjective and an upcoming predictable noun. Their results revealed an ERP effect locked to the adjective in the mismatching condition, which showed participants' sensitivity to the unfolding syntactic violations between the adjective and their predictions. Taken together, these studies provide instances of anticipatory processing in both syntactic and semantic domains that show prediction occurs at differing linguistic levels.

These studies treat the speaker as being perfectly fluent in production. However, this does not reflect the ecological reality of everyday spoken language where disfluency affects approximately 6 in 100 words (Fox Tree, 1995). Disfluency has been shown to impact upon language comprehension, influencing the parsing of garden-path sentences (Bailey & Ferreira, 2003), attenuation of context dependent word integration (Corley et al., 2007) and speeding up of word recognition (Corley & Hartsuiker, 2011). In the longer term, listeners show an increased likelihood of remembering words that appear immediately after encountering disfluency (Collard et al., 2008; Corley et al., 2007).

It follows that there have been attempts to categorise and understand the underlying mechanisms that are responsible for these disfluency effects seen during comprehension. Different models of disfluency processing have been proposed to explain these effects, with the current study concerned with trying to differentiate between two: the predictional and the attentional accounts.

Disfluency has been shown to modulate predictive processing, with a clear effect in the literature being that upon encountering the filled pause, *uh*, listeners show a bias for unknown or discourse new referents (Arnold et al., 2007, 2004; Heller et al., 2014). Bosker, Quené, Sanders and de Jong (2014) showed that similar results held following an *um*, with listeners demonstrating a preference towards low-frequency referents, over high-frequency objects. This forms the basis of the predictional standpoint for disfluency processing (e.g., Arnold et al., 2007; 2004; Heller et al., 2014) that suggests that upon encountering disfluency, a listener infers the speaker to be experiencing difficulty. This difficulty can be driven by the situation of the speaker, with an increased tendency to be disfluent when they are experiencing cognitive load (Bortfeld et al., 2001; Brennan & Schober, 2001) or in the face of increased difficulty in lexical retrieval, for example when trying to produce a word that is contextually unpredictable or low frequency words (Beattie & Butterworth,

1979). Listeners use this knowledge to build up patterns of disfluency distribution information that inform their expectations of upcoming content for a speaker.

The reliance on this distributional information or speaker modelling is flexible and can be modulated by other knowledge that influences the cognitive representation of the speaker, whether this is that they have difficulty naming objects (Arnold et al., 2007), there are multiple speakers each with a different set of discourse new and old objects (Barr & Seyfeddinipur, 2010) or that the speaker is non-native and may have a variable pattern of disfluency (Bosker et al., 2014). Heller et al. (2014) further demonstrated the flexibility of the online disfluency processing mechanism in response to situational and speaker specific contextual information. They showed that instead of listeners directly associating disfluency with certain properties of objects, for example, a lack of conventional name, they used situation-specific inferences to guide their predictions for upcoming referents. However, they also revealed limitations to these inferences, as listeners showed a lack of sensitivity to assumed speaker knowledge of referent names when it diverged from their own experience of the names, in contrast to the results seen in Barr and Seyfeddinipur (2010). In summary, the predictional account of disfluency processing relies on a probabilistic attribution of speaker difficulty, coupled with situational and speaker specific knowledge to infer the cognitive state of the speaker and uses this information to update expectancies for upcoming content.

A competing attentional account has been proposed to explain comprehension effects seen which suggest that following disfluency listeners employ heightened attentional resources. Fox Tree (2001) tested the impact of filled pauses, *uh* and *um*, on the time taken for word identification. The results showed that following *uh*, participants were quicker to identify a target word than in the related condition that featured a silent pause of the same duration. In contrast, participants did not take less time to respond following *um* than following a silent pause. The theory offered by Fox Tree for this lack of facilitation effect is that *uh* and *um* are different words, with *um* thought to represent a longer upcoming delay in speech. Fox Tree suggests

that orienting attention would be impractical when the time course of the resumption of speech is unknown. However, Corley and Stewart (2008) proposed an alternative explanation for the lack of effect seen for *um* in Fox Tree's study (2001), namely that the duration of the remaining silent pause from the removed *um* represents a delay that extends beyond a normal gap in fluent speech, as it is notably longer than for either the filled or silent pause *uh* condition. Therefore, the silent pause in the *um* condition could have been comprehended as disfluent or processed in a manner divergent from typical fluent speech.

Further support for an attentional mechanism in disfluency processing was demonstrated by Collard, Corley, MacGregor, & Donaldson's (2008) ERP experiment that showed that novel stimuli presented after a disfluency elicited a decreased amplitude in the attention associated brain component (P300). The reduction seen in this component suggests that participants were already attending to the incoming speech, providing support for the viewpoint that following a disfluency, listeners are orienting their attention to the upcoming content and this heightened attention is responsible for facilitation effects seen following filled pauses (e.g., Brennan & Schober, 2001; Fox Tree, 2001). A possible reason behind this facilitation is that the disfluency causes listeners to abandon predictional processes and rely on bottom-up information, the incoming speech signal, to resolve the comprehension difficulty posed by the interruption to the speech, whilst the increased attentional resources allow quicker recognition of following linguistic content. It also possible that these two accounts are not mutually exclusive and this is an idea we come back to in the general discussion at the end of the thesis.

In the present study, we aimed to distinguish between the predictional and attentional disfluency processing accounts by employing a visual world paradigm to investigate directly the underlying processing during comprehension. Participants heard an utterance whilst being eye-tracked viewing a scene. After the utterance finished they were asked to click on the referent (target item) heard. Half

of the time, the sentence stimuli heard featured a disfluent production to measure the impact of encountering disfluency on the eye-movement behaviour. The current study utilised a 2x2 plausibility vs accessibility design to examine the comprehension processes taking place after disfluency is heard. We manipulated the accessibility of a target word from the preceding sentence context by employing high and low cloze probability completions (e.g., Taylor, 1953). Cloze style completions have been used previously as an effective measure of contextual constraint (e.g., Kutas & Hillyard, 1984). Plausibility of the target word within the context of the preceding sentential stem was also measured. Each scene included 4 objects that had a set role to the sentential stem [Example-" *The vet was sad to do so but he had to put down the local family's trusty…*"]: 1 semantically related Highly Accessible item (High Cloze [*Dog*]); 1 semantically related Low Accessibility item (Low Cloze [*cat*]); 1 Semantically Related implausible item (SR [pills]) and 1 Semantically unrelated distractor item (SU [vest]).

Crucially, the predictional and attentional accounts differ in how they would predict comprehenders' patterns of looking towards objects in a visual scene following a disfluency and before the onset of the target word signals the correct target. In the fluent condition the accounts make the same prediction: participants are likely to anticipate the picture they will have to click due to the constraining context and this will lead to fixations on the predicted (HC) object, as this the most plausible item to finish the sentence. However in the disfluent condition the accounts predictions diverge: If disfluency is signalling to listeners that the speaker is experiencing difficulty as proposed in the predictional account, then the listener would predict the upcoming object to be harder to access for the speaker and an increased proportion of looks to the competitor (LC) picture would be expected, as it the only other plausible object in the scene. If, as proposed in the attentional account, expectations of the upcoming content cease following a disfluency, then this will cause them to abandon or attenuate predictions that the sentence will end with the 'predicted' HC item. Instead, we would expect the sentence context to exert a weaker effect and this would increase fixations on *both* the competitor (LC) item

and the semantically related (SR) item.

3.3 Method

3.3.1 Experimental Scenes

The predictions made are based on the objects pictured in a scene having a defined set of relationships to the corresponding sentential contexts heard in that trial; these relationships are described here. An experimental scene always consisted of 4 pictures, with each picture having a set relationship to the sentential context (Role): High Cloze (HC) picture, highly predictable and plausible in the context; Low Cloze (LC) picture, less predictable but equally as plausible as the HC items; Semantically Related (SR) picture, unpredictable and implausible in the context but semantically related to the sentential context and Semantically Unrelated (SU) picture, a distractor that was unrelated to the sentential context. Figure 4.1 shows an example scene taken from the current study. The target word heard for a trial always matched either the HC or LC picture from the corresponding scene.

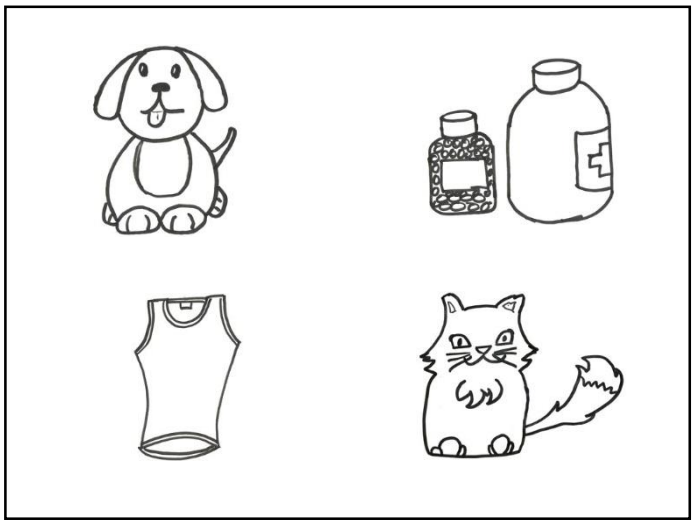
3.4 Norming Studies

A number of pre-tests were undertaken to gauge the suitability of materials. First, we present a cloze task to test potential target words and following that, a plausibility norming study. These norming studies are described in detail below.

3.4.1 Cloze-Task

The first pre-test undertaken was a cloze task (e.g., Taylor, 1953). We asked participants to read a sentence and then fill in the sentence final blank. This pre-test was concerned with creating sentential contexts for the main experiment that were constraining enough to create only a small number of words that could complete the sentence. Ideally, only a single completion. A cloze-test gauges how participants would complete a sentence and the results provide a proven measure of the strength

of the constraint that our test sentential contexts created (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kutas & Hillyard, 1984).



"The vet was sad to do so but he had to put down the local family's trusty…".

Figure 3.1- An example experimental scene. High Cloze (HC) picture- DOG, Low Cloze (LC) picture-CAT, Semantically Related (SR) picture-MEDICINE, Semantically Unrelated (SU) picture-VEST.

A total of 32 undergraduate students from the Psychology community at the University of Edinburgh completed the task. Participants were recruited in return for course credit; this method was used for all norming studies. Five further participants were excluded for not being native English speakers. Participants saw 60 test sentence fragments that followed the same syntactic structure throughout: each sentence started with an agent which performs an action to an object. For example: "*The vet was sad to do so but he had to put down the local family's trusty…*" (as seen in Figure 4.1). The participants were instructed, "Please fill in the sentences in a natural manner, a single word should be enough. There are no right and wrong answers. So don't think about it too hard...". We encouraged participants to answer instinctively so that we elicited their natural answer. The cloze test was run online using an online survey tool (Bristol Online Surveys at www.survey.ed.ac.uk).

Participants had to complete the sentence by typing a word into a box following the sentential context. The study took approximately 20 minutes to complete.

Those responses that were produced by over half of participants were labelled as HC. We initially selected 40 HC completions, to be tested further in the Plausibility Norming study. The 40 LC endings were not directly taken from participants' responses. Instead, they were selected by the experimenter for their suitability in the paradigm in a LC role. However, some of the chosen LC endings matched responses given by a small number of participants. From these 40 HC/LC endings, 28 of each ending type were selected for inclusion in the main experiment after the Plausibility norming study described below. There was a large difference of 61% between the mean response rates for HC and LC endings chosen for the main experiment: 68% for HC and 7% for LC. This was shown to be significant ($t=18.26$, $df=54$, $p<0.001$).

3.4.2 Plausibility Norming

After the HC and LC items had been selected, plausibility tests of the potential audio stimuli followed to assess the contextual fit of the items with the sentence stems. This was crucial as our predictions were based on experimental scenes containing HC and LC items with equal or close to equal plausibility. The pre-test asked participants to read sentences and then, "Rate the plausibility of a final word, for a number of sentences. Please think about the sentences in a natural manner, and assign a rating." The sentences were presented with the target word in bold, for example:

"*The vet was sad to do so but he had to put down the local family's trusty* ***dog***."

There were two types of trial: Experimental and Filler. The Experimental trials were made of a sentence context and a target. The same 40 sentential contexts used in the previous Cloze pre-test were used again**.** Each sentential context was presented with 4 target types: the HC item selected from Pre-Test 1; the LC item selected in the previous study; and two extra variants of the LC condition. These additional LC targets were included to maximise the chances of finding a LC item which matched

the HC item for plausibility. The filler trials exhibited a range of plausibility; for example, a plausible version was, "The dog was barking loudly from inside the red kennel." An example of the less or implausible items was, "The Koala had been invited to the teddy bear's picnic. He was excited and had bought a new suit." The Fillers followed the same sentence structure as the experimental items, so there was no obvious differences to differentiate the fillers. To counterbalance the different HC and LC items, the second plausibility norming study was sub-divided into four lists with each forming a separate online pre-test. Participants only ever saw one list. This pre-test was also run online using the same online survey tool as the previous pre-test (Bristol Online Surveys at www.survey.ed.ac.uk). Participants had to select a rating, by clicking on a button corresponding to the appropriate rating. A standard Likert scale (Likert, 1932) of 1-7 was used for rating, where 1 was highly implausible and highly plausible was 7.

Participants saw 83 test sentences: 40 experimental items and 43 filler items. Each participant saw 10 HC targets and 30 LC targets. The filler items were the same for all 4 conditions. The study took approximately 20 minutes to complete. A total of 42 undergraduate students from the Psychology community at the University of Edinburgh participated in the final online plausibility test to gain course credit (List 1=10; List 2= 15; List 3= 8; List 4=9). The number of participants for each list was not balanced due to a number of self-identified non-native speakers performing the pre-test, leading to the exclusion of their data. The target number of participants for each list was 10 people. However, list 2 was erroneously uploaded twice, explaining the increased number of participants (15) that completed this List. To select the final LC item from the 3 variants for each experimental scene, we chose the word that scored the highest plausibility rating. The selected LC items only were analysed. The results from this final norming study showed a small difference between the mean HC (5.6) and LC (5.22) plausibility ratings. However, it was still at a significant level ($t=2.1$, $df = 54$, $p=0.041$).

Theoretically we wanted HC and LC items to be rated equally plausible for the same sentential context. However, the difference in cloze completion rates demonstrate that the sentential contexts are more constraining towards the HC item, as seen previously in Federmeier et al. (2007): Participants are frequently more likely to choose a HC item to finish a sentence over a LC item, the defining property of cloze probability. The cloze probability of a sentence final target word has been shown to affect processing, with facilitation for the 'best completion' and graded facilitation for lower cloze probability words, even with semantic overlap between completions (e.g., Kutas & Hillyard, 1984). It was clear then that for the current materials there was a stronger associative link between the sentential context and the HC referents over the related LC referents, for example- 'miner' and 'coal'/'jewels' respectively. This is supported by findings from Federmeier et al. (2007) that showed that when unexpected (defined by low cloze probability) but plausible words completed highly-constraining sentences there was a late-occurring (500-900ms) component that is sensitive to the mismatch between the expectancy generated by the context and the unexpected target word. This result shows that participants are sensitive to the difference in strength of the sentential constraint between HC and LC items. For the current study, this difference in sentence constraint was likely driven by the frequency of co-occurrence between the context and HC/LC endings, leading to increased expectancy of one completion: Participants were more used to hearing 'miner' with 'coal' than 'jewels'. It was perfectly plausible that a 'miner' could mine 'jewels' but the prototypical answer, as reflected in the cloze task, was 'coal'. Therefore, the strength of expectancy between 'miner' and 'coal' is causing participants to rate anything that is not 'coal' as less plausible. To put this another way, although, semantically and contextually equally plausible, the highly constraining context creates a 'best completion' that creates a greater expectancy and participants recognise this as the best fit. Although others endings are plausible this 'best fit' item has special status and hence, they view other items as less plausible. This phenomenon occurs even though the HC and LC may be semantically related (Federmeier et al., 2007). We were unable to

successfully match the plausibility of the HC and LC targets; a reliable difference in plausibility will therefore have to be taken into account when interpreting the findings.

3.5 Experiment 1: Visual World

3.5.1 Participants

A total of 32 students from the University of Edinburgh participated for a reward of £5 upon successful completion of the experiment. Participants self-reported that they were native speakers of English and had normal or corrected to normal vision. Participants who had taken any of the pre-tests were excluded from taking part in the main experiment.

3.5.2 Design and Materials

Each trial contained a contextual sentence stem being presented with an experimental scene, followed by a target item. The current study included experimental trials and filler trials. Both trial types could be either fluent or disfluent. Each experimental scene contained four pictures, as described above: a High Cloze picture (HC); a Low Cloze picture (LC); a Semantically Related (SR) picture and Semantically Unrelated (SU) picture. The scenes that accompanied filler trials featured 1 target picture and 3 unrelated, distractor pictures. All pictures were hand drawn by author and then scanned. All pictures were presented in black and white and at a size of 380x300 pixels, as part of a 1024x768 pixel screen. Participants saw the pictures used for experimental scenes once. Filler trial pictures were again seen only once, aside from those used in the practice trials. The pictures used in the practice trials were seen again in filler trials in the main experiment. An example sentential context and experimental scene can be seen in Figure 4.1 above.

The contextual sentence stems followed the same structure throughout all experimental trials; Starting with an agent which performs an action to an object. For example: *"The vet was sad to do so but he had to put down the local family's trusty…"*

Filler trials followed a loosely similar syntactic structure, with each starting with an agent and ending in a noun. However, they were designed to be less constraining and predictable: *"Amy wanted to play again after watching the final on television but she couldn't find her racket."* Both types of contextual sentence stems were presented in fluent and disfluent versions. In disfluent experimental trials the filled pause (*uh*) was always presented in the same position in the sentence: prior to the sentence-final target item. The disfluent version of our example was, "*The vet was sad to do so but he had to put down the local family's trusty UH…*"The average duration of the filled pause, (*uh*), used for the experimental trials was 437ms (SD=87ms). The disfluent fillers always used the same disfluency, (*uh*). However, the position of the disfluency differed; they appeared across the sentence in pre-nominal locations. An example disfluent filler was, *"The old man from UH Chester had a collection of ten thousand stamps."* The disfluency position was varied so as to de-emphasize the repeated position of the disfluency in experimental trials.

A challenge with the investigation of disfluency is exerting necessary experimental control whilst maintaining a production that is close to a natural manner. In the current study, we aimed to keep the recording as naturalistic as possible which resulted in disfluent sentential contexts being an average of 936ms longer than the fluent contexts. The time course differences were due to notable differences in the pause length between the end of the sentential context and the target for the disfluent contexts with an average 341ms (SD= 107ms), an extra average 301ms pause over the fluent variants. Additional silent pause duration following filled pauses were not unique to the current materials and had been reported previously (e.g., Clark & Fox Tree, 2002; Fox Tree, 2001). With the addition of the disfluency duration (437ms) this accounts for 778ms of the extra duration. The 158ms of remaining time was differences in delivery between fluent and disfluent contexts. The extra silent pause duration was not excised, as shortening this pause led the experimenter to judge the sentence contexts as having a less natural sounding delivery.

The target item in experimental trials was always either the HC or LC referent related to the sentential context, as described in the pre-test above. Half of the experimental trials ended with a HC target and half with a LC target.

Condition	***Fluency of Sentential Context***	***Target***
1	Fluent	High Cloze Predicted
2	Fluent	Low Cloze Competitor
3	Disfluent	High Cloze Predicted
4	Disfluent	Low Cloze Competitor

Table 3.1: The Experimental Conditions.

1a) *List1*: "The vet was sad to do so but he had to put down the local family's trusty DOG" [Condition 1: Fluent/HC].

1b) *List 2*: "The vet was sad to do so but he had to put down the local family's trusty CAT" [Condition 2: Fluent/LC].

1c) *List 3*: "The vet was sad to do so but he had to put down the local family's trusty ('UH') DOG" [Condition 3: Disfluent/HC].

1d) *List 4*: "The vet was sad to do so but he had to put down the local family's trusty ('UH') CAT" [Condition 4: Disfluent/LC].

The experiment consisted of 48 trials: 28 experimental items and 20 filler items. There were 4 lists; all lists contained the same 28 experimental sentential contexts but a different condition was presented in each list. Using the experimental trial depicted in Figure 3.1 above as an example, Table 3.1 shows the utterances for each

condition that the sentential contexts could be presented in. These examples demonstrate how each list contained a different condition variant of each trial; the scene would be identical for each participant but the utterance and target picture to be selected would differ across lists (1a-d). Each list contained 7 sentential contexts from each of the conditions. This meant equal numbers of fluent/disfluent presentations and HC/LC targets in each list. The same set of 20 fillers were used for each list. Half of the filler trials were presented in a fluent manner and half containing disfluencies. The experiment was presented to the participants in 2 blocks of 24 trials with an optional break in between. Trials were presented in a random mixed order. Each participant saw only one block. The design was 2 (fluent vs. disfluent) X 2 (Predicted or Competitor Target).

The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the experimenter present. A native British English speaker was recorded producing all of the materials. The speaker completed the recording of all materials in one session. The speaker was instructed to produce a disfluency of a natural length. Sentential contexts and target items were recorded separately and repeated until the experimenter judged a delivery approximating natural speech was achieved. The sentential contexts were always produced with the token “pen” as the final referent. This minimised the effects of co-articulation and prosody between the sentence and target, as otherwise this could have provided the participant with clues as to which item was going to be named before the onset of the target. The sentential context was kept as recorded with only the final token excised. To remove the 'pen' token, we viewed the sentence contexts as a sound wave; cutting immediately before the start of the distinctive shape of the plosive 'p'. The target items were recorded using a number of neutral sentence place holders to minimise list effects on the pronunciation of the target and keep the production replicating natural spoken language as much as possible. Each sentence had a pre-target word-final plosive phoneme to make it easy to judge where the place holder ended. The sentences were viewed as sound waves and the context was then excised

immediately after this phoneme to leave just the target. All recordings were saved in a mono 48kHz .wav format.

3.5.3 Apparatus and Procedure

The visual and audio stimuli were presented using 'Experiment builder' software (version 1.10.165, SR Research, 2012) on a PC and a 15 inch monitor set at a 1024x768 resolution. The eye-tracker used was a table mounted SR Research Eye-link 1000, which sampled eye-location at a rate of 500Hz.

After reading an information sheet and filling in a consent form, participants were seated at the eye-tracker, with their head in a stand to minimise head movement, looking at a computer screen. Participants then read through the practice instructions presented onscreen. The practice instructions gave a brief description of the task that followed, including a reminder to keep their heads still. Following this, they performed 4 practice trials, which did not vary across participants. These practice trials comprised of 4 filler trials which were repeated in the main experiment and were designed as a familiarisation phase. Experimental trials were not repeated as this would have led to repetition which could have affected responses to the replicated trials. The practice trials followed the exact same structure as the main experiment, aside from the lack of calibration and need to trigger a fixation check at the beginning of each trial as the participants were not being eye-tracked during the practice trials. However, the fixation check screen was still included so participants followed the same procedure as in the main experiment. Participants did not have to trigger the fixation check, instead they clicked to move to the next scene. The experimental scene was presented for a second before the utterance began playing. Upon completion of the target, an orange circle representing the mouse cursor became visible in the middle of the screen. Then participants had to move the mouse and click on the referent heard in the auditory stimulus. Once the mouse-click had been registered, the fixation check screen for the next trial appeared, indicating the beginning of the subsequent trial. After the practice trials, the participants were always given a chance to ask

questions and the experimenter checked that they felt familiar with the structure of each trial.

After this the participants began the main data collection phase of the experiment. This section began with instructions for the main experiment. The instructions stated that participants would be calibrated before the trials began and how to trigger the fixation check. A brief description of the task followed, including a reminder to keep their heads still. They were also informed that there would be an optional break during the experiment. Following the instructions and after each block participants completed a 9-point SR Research calibration routine. This calibration routine could be accessed after each trial if the participant moved from their position or presented any other behaviour that potentially reduced the accuracy of that calibration. Then the trials began. Each trial started with a fixation check; a small grey box (25 X 25 pixels) located in the centre of the screen which the participants had to trigger, by moving their gaze to within the box, to proceed to the experimental scene. The rest of the trial was as described above for the Practice trials. Participants then saw 48 trials in two blocks of 24. The study lasted approximately 60 minutes.

3.5.4 Measures

The current study examined the following measures:
-*Eye-movement behaviour over time*: The primary measure was the proportion of fixations on each picture during the experimental trial period. The eye-tracker sampled the location of the fixation on the screen every 2ms.
-*Correct picture selection*: the picture clicked on by the participants for each experimental trial was recorded.

3.6 Analysis

We analysed participants' eye movements and the image that they clicked on. Eye movement data was analysed as follows. Rectangular interest areas of identical size (380 x 300 pixels) were created to capture looks to each of the 4 images presented in the scene. The interest areas did not cover the whole scene but were equally spaced from the centre point and the margins of the scene. The remaining area of the scene was coded as background. Fixations that fell outside of these 4 interest areas were not counted in the interest area fixations and during that time period there were no fixations on any interest area. Although all participants were successfully calibrated on the eye-tracker before beginning the experiment, three participants' data had to be excluded as the eye tracker did not record their data correctly. Of the remaining data, we eliminated trials where the participant clicked on the wrong item. This occurred on 0.6% of trials.

Certain time points were particularly important and were marked in each trial: (a) the sentence onset (b) target onset and (c) the response time of the participant's clicking on an object. The analyses were based around the coded target onset point in the Eyelink output files. Thus we had a full record of eye movements relative to these points of the speech. Fixations which started before the boundaries of the time periods and continued into the set time period were included in the analyses. Those fixations which originated in the set time period but extended past the final boundary were not included. This process was automated across all trials by the software and not controlled by the experimenter. The time taken by blinks was recorded and added to the fixation time in a given interest area. Our primary focus was the pattern of eye movements and the objects being fixated over two set time periods: 1000ms before until Target Onset (Pre-TO) and 600-1000ms after the target onset (Post-TO). Each time period was analysed separately.

Firstly, we analysed the proportion of fixations for each role by fluency and target conditions from the onset of the target word and the 1000ms preceding it (Pre-TO). The predictions for this study relied on the potential differences in the patterns of looking between fluency and target conditions. This 1000ms period provided the

range of time in which we would expect the fluent and disfluent conditions to lead to divergent behaviour. On average, a disfluent sentential context was 936ms longer than a fluent context, see Table 3.2.

Table 3.2: Duration of sentential contexts and targets by condition in ms (Standard deviations in brackets).

CONDITION	*SENTENCE PLACE-HOLDER*	*DISFLUENCY (AVERAGE)*	*AVERAGE PRE-TARGET GAP*	*TOTAL AVERAGE DURATION*	*TARGET*	*AVERAGE DURATION*
1	Fluent	-	40 (22)	4560 (1242)	Predicted (HC)	595 (150)
2	Fluent	-	40 (22)	4560 (1242)	Competitor (LC)	685 (177)
3	Disfluent	437 (87)	341 (107)	5496 (1254)	Predicted (HC)	595 (150)
4	Disfluent	437 (87)	341 (107)	5496 (1254)	Competitor (LC)	685 (177)

The disfluent context is longer because of the increased pre-target gap and the inclusion of a disfluency of an average 437ms duration. Therefore in the disfluent condition this time period would cover part or all of the disfluency and the following pre-target gap and in items with a shorter disfluency some of the preceding sentential context will be included. The standard assumption of the average time within visual world studies to initiate and execute a saccade is up to 200ms (Matin, Shao & Boff, 1993; Altmann & Kamide, 2004). Recently this assumption has been revised down to around 100ms (Altmann, 2011) but this remains contentious (Salverda et al., 2014). However, in this 1000ms time window by either assumed saccade execution time, it would be reasonable to plan and launch a number of saccades upon encountering the disfluency, allowing sufficient time to capture possible changes in the fixation behaviour of participants.

This same time period (Pre-TO) in the fluent condition would incorporate the final determiner/adjective and preceding sentential context, dependant on items. Until

the onset of disfluency, the participants would hear contextually identical stimuli in both fluent and disfluent sentences. At this point, participants would not have heard any clues to the identity of the target and, hence, the picture that they would have to select. All items were recorded with same target, so there could be no co-articulation clues for the upcoming target, so this could not have influenced the patterns of looks. Therefore, you would not expect any differences between the predicted (HC) and competitor (LC) conditions in this time period. This data is plotted in Figure 3.4-3.7.

The second time period we analysed was 600-1000ms after target onset (Post-TO). Although the predictions we present are only valid in the time period until target onset, this second time period was useful as a check that the participant's fixation behaviour aligned with the target heard and lend any differences in the earlier time period greater validity. During this period, across all conditions, listeners will have heard the target and would be expected to fixate on the picture that they have to select to complete a trial most often. This pattern of results seen during this time period is shown in Figures 4.4-4.7.

Participants process referring expressions incrementally and constrain looks towards referents which are still possible following the information heard until that point (Sedivy, Tanenhaus, Chambers, & Carlson, 1999). On encountering a visual scene participants will initiate eye-movements to a picture once it becomes the unique referent (Eberhard et al., 1995; Sedivy et al., 1999). It is uncontentious then that upon hearing the target words, participants will fixate on that referent within the scene. For the current experiment, as the average duration for all targets was 640ms; this meant that even if participants did not plan a saccade until each target was fully produced, the additional 200ms to execute it would fall well within this time period for the vast majority of the items.

There were 2 targets whose durations exceeded the upper limit, 1000ms, of the Post-TO period. However during this period participants will still have narrowed down to a unique referent and are likely to have started anticipatory looks to this picture.

Overall, this was only a very small number of targets (7%) and we would still predict differences between predicted (HC) and competitor (LC) targets for these targets. It was likely that participants can predict the picture they will have to select much earlier post target onset, as previously it has been shown that participants are sensitive to online phonetic information much earlier than the offset a word. Participants begin to fixate on a referent over a competitor that begins with a contrasting phoneme in a scene after hearing the onset of the target word (Salverda et al., 2007). If referents begin with the same phoneme then participants process the phonemic information incrementally and have to wait until following phonemes to select a unique referent (Sajin & Connine, 2014). In 16 trials (14%), there was a phonological competitor present in the scene. However, during this period participants would have narrowed down to a unique referent and were likely to have started anticipatory looks to this picture. Therefore, we would still predict differences between predicted (HC) and competitor (LC) targets in these trials.

It follows that from 200ms post target onset onwards participants could realistically be expected to fixate on the referent heard. Although these fixations to the target may have occurred before the 600-1000ms time period (Post TO) being measured here, the mouse pointer the participants used to select an item did not become active until after target was completed. Therefore it was unlikely that participants would look away from that referent until they had selected an object and completed the task. It would also take time to plan and execute the motor movement of moving the pointer to the correct target during which time the participant is likely to be fixated on the goal picture to ensure correct selection is made.

The 400ms duration of the 600ms-1000ms period (Post-TO) was shorter than the previous 1000ms duration of the pre-TO period. This duration was chosen because there was a specific window during which the vast majority of targets would end and the fixations resulting from processing this information were most likely to be at the target during this period and 400ms would allow a minimum of 2 saccades to take place during the period. The same pattern of results is seen in a matching

period of duration 1000ms from 400ms-1400ms post target onset. However, by the end of this period, most targets will have ended and towards the end of this period it will be unduly influenced by a smaller number of trials, whereas the shorter 400ms period will feature all trials. Therefore, for the current study within this time period participants will have narrowed down the set of available referents to the one that they heard for almost all targets and we would expect there to be differences between the predicted (HC) and competitor (LC) conditions due to this.

The data and the subsequent analyses had to be separated by role and are presented in this order: HC, LC and SR. SU is not analysed as it does not have consequences for our predictions. We could not collapse across conditions because each picture and the role it had were not independent, as they had co-occurred with all other roles within a scene. This meant that when participants were fixating on any picture, and hence role, at a given time they could not fixate on any of the other pictures: If a participant was fixating on the HC picture, they could not be also be fixating on any of the remaining LC, SR or SU pictures. As our dependent variable was binomial we arcsine square root transformed the proportions of fixations before running ANOVA analyses (e.g., Altmann & Kamide, 1999). The predictors we use in the analyses are fluency condition (Fluent/Disfluent) and Target (Predicted (HC)/ Competitor (LC)) which were within subjects and items. All analyses were carried out in R (R Development Core Team, 2014).

3.7 Results

In 4 scenes (14%), there was a phonological competitor to the target present in the scene. In these trials, ambiguity about target selection was extended for a marginally longer time than when no phonological competitors were present. In the remaining trials, the target onset uniquely identified the referent in the scene. However, if a phonological competitor were present, a unique referent could not be selected until a greater amount of the target had been heard because more than one referent in the

scene began with the same initial phoneme. However, our predictions relate to eye-movement behaviour before target onset. Therefore, these competitor trials would still speak to our predictions, as having a phonological competitor referent in the scene does not affect participants' expectations pre-target. The predictive behaviours are hypothesised to be based on semantic and plausibility knowledge.

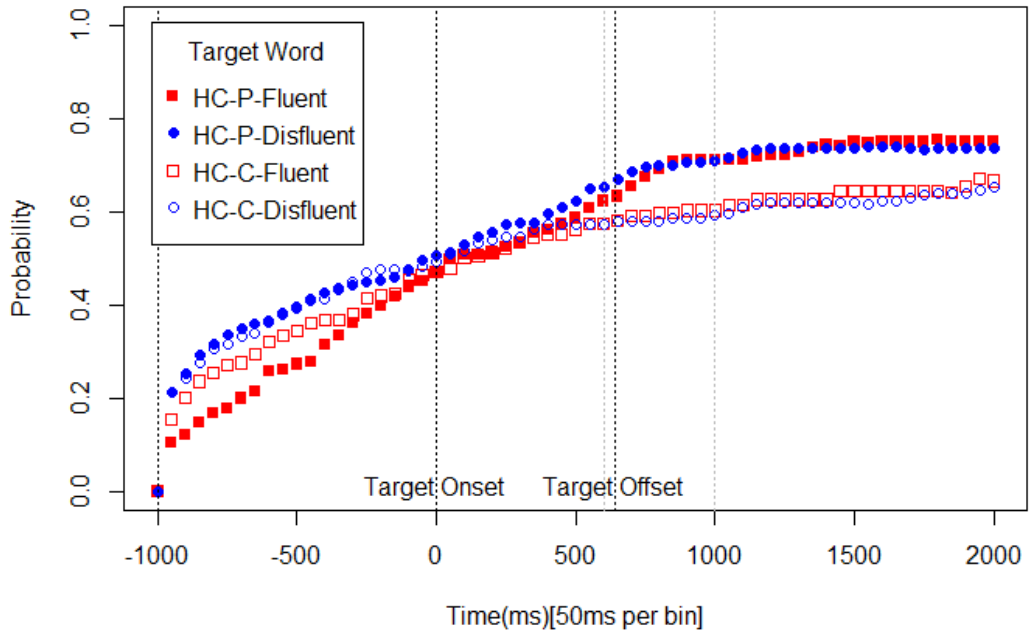


Figure 3.2 - The cumulative probability of fixating on the HC Picture as a function of Fluency of Sentential Context (Fluent vs. Disfluent) and Target Heard (Predicted-P vs Competitor- C). Target Onset and Offset (average) are marked. Analysis periods marked: -1000 to Target Onset; 600-1000ms after Target Onset.

We calculated the cumulative probability of looks to each of the picture roles (HC, LC, SR, SU) with each of the fluent and disfluent sentence contexts and for both targets heard (Predicted [HC], Competitor [LC]) for a set period from 1000ms before until 2000ms following the target onset in 50-ms intervals as in Altmann and Kamide (1999). This data is plotted in Figures 4.2-4.5. These figures provide visualisation of the patterns of fixation behaviour that unfold over the time periods

being analysed.

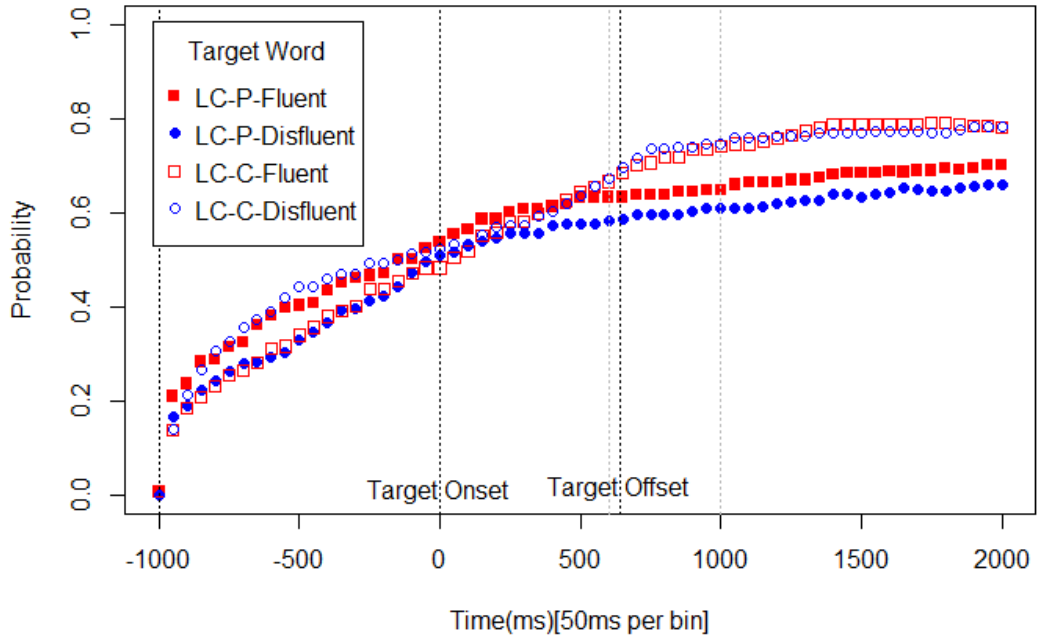


Figure 3.3 - The cumulative probability of fixating on the LC Picture as a function of Fluency of Sentential Context (Fluent vs. Disfluent) and Target Heard (Predicted-P vs Competitor- C). Target Onset and Offset (average) are marked. Analysis periods marked: -1000 to Target Onset; 600-1000ms after Target Onset.

The analyses were subset by fixations to each of the three theoretically relevant images, as described above (HC, LC, and SR). First, we report the results for the -1000ms to Target Onset (Pre-TO) time period described above. The proportions of fixations broken down by role, fluency and target can be seen in Figure 3.6. A two-way ANOVA on the arcsine-square root transformed proportions for the HC data with fluency and target as predictors revealed a main effect of participants making more fixations on a HC picture following a disfluent context for both types of targets, (F1 (1,28)= 5.78, MSE= 0.15, p=0.02; F2 (1,27)= 6.22, MSE= 0.18, p=0.02). Participants had an average 0.06 increase in fixation proportion on the HC picture in a disfluent context collapsed across Target type (means- Fluent: 0.27, Disfluent: 0.33). There was no difference between Target conditions or any interaction effects seen for the HC data (all Fs <2).

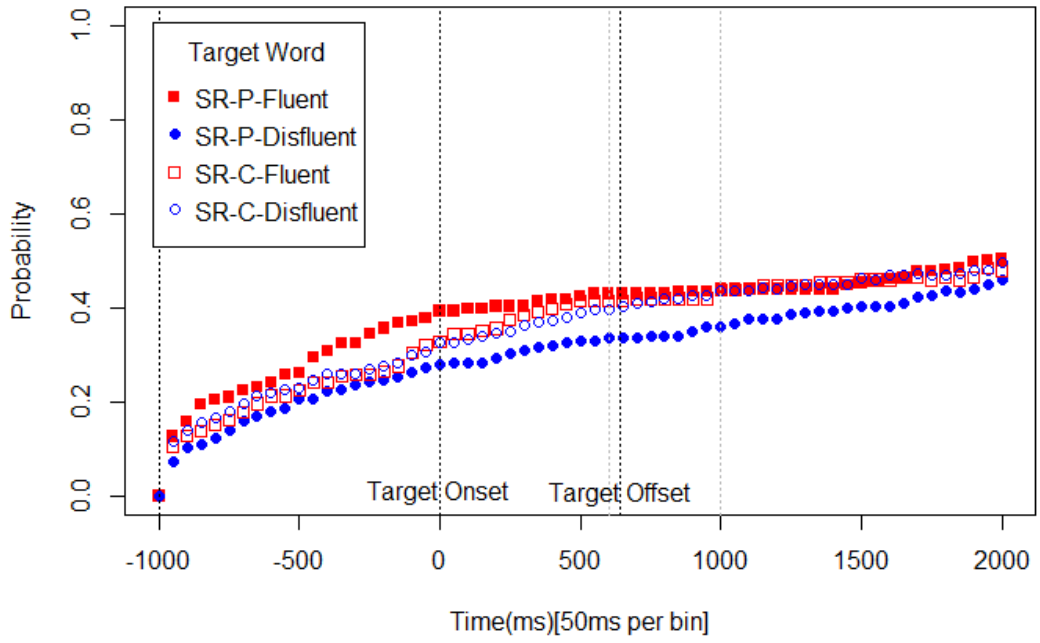


Figure 3.4- The cumulative probability of fixating on the SR Picture as a function of Fluency of Sentential Context (Fluent vs. Disfluent) and Target Heard (Predicted-P vs Competitor- C). Target Onset and Offset (average) are marked. Analysis periods marked: -1000 to Target Onset; 600-1000ms after Target Onset.

The same ANOVA ran on the arcsine-square root transformed proportions for the LC data showed no main or interaction effect for either Disfluency or Target (all Fs <1). The average proportion of fixations following a fluent sentence context for the LC data was marginally higher (0.29) than the average for the equivalent HC data (0.27). However, the average proportion of fixations following a disfluent context for the LC came out as slightly less (0.31) compared to the HC data (0.33). For the SR data, participants had an average 0.05 increase in fixation proportion on the SR picture following a fluent context (means: Fluent= 23; Disfluent=0.18); this main effect, seen by both participants and items, was in the opposite direction to the disfluency effect seen for the HC picture (F1 (1, 28) = 6.36, MSE= 0.17, p=0.018; F2 (1, 27) = 6.08, MSE= 0.13, p=0.02). There were no other main or interaction effects for the

SR data (all Fs <1). The SU data is not reported due to it not having any significance towards distinguishing between our predictions.

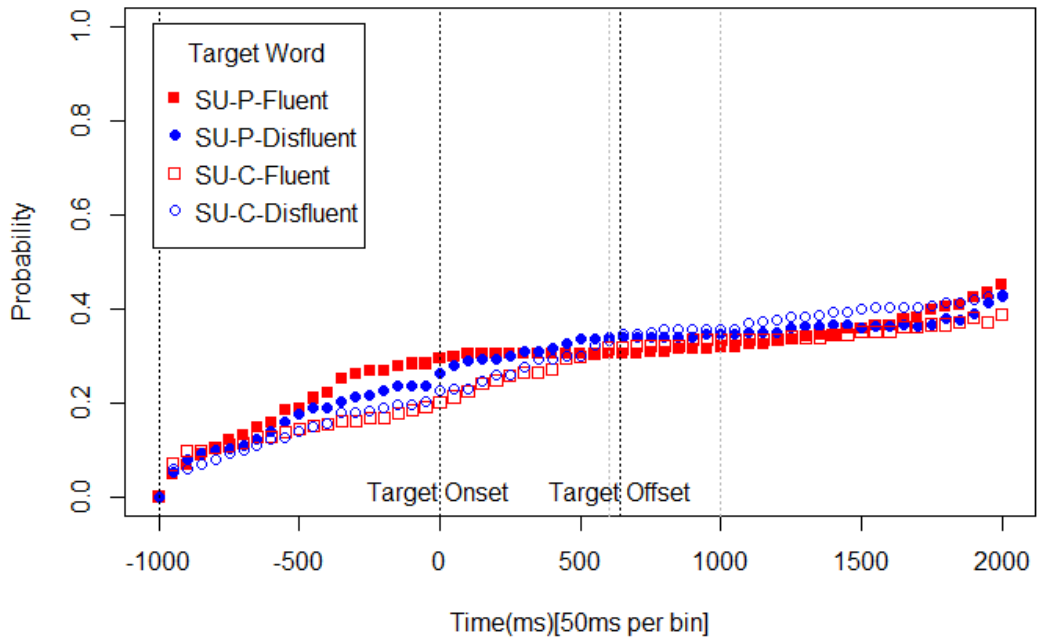


Figure 3.5 - The cumulative probability of fixating on the SU Picture as a function of Fluency of Sentential Context (Fluent vs. Disfluent) and Target Heard (Predicted-P vs Competitor- C). Target Onset and Offset (average) are marked. Analysis periods marked: -1000 to Target Onset; 600-1000ms after Target Onset.

The second time period analysed was 600-1000ms after target onset (Post-TO). There was a different pattern of results for target during this period: The proportions of fixations broken down by role (HC, LC), fluency and target and this can be seen in Figure 3.7. A two-way ANOVA on the arcsine-square root transformed proportions for the HC data with fluency and target as predictors showed a massive increase of 0.42 in fixation proportion for Predicted (HC) target (mean: 0.61) compared to Competitor (LC) targets (mean: 0.19) by participants. This was realised as highly robust main effect for Target by both subject and item analyses ($F1 (1, 28) = 174.6$, $MSE= 6.45$, $p<0.001$; $F2 (1, 27) = 191.1$, $MSE= 6.23$, $p<0.001$). There was no other main or interaction effects for the HC data (all Fs <1).

The same ANOVA ran on the arcsine-square root transformed proportions for the LC data showed a main effect of Target, by subject and item, in the reverse direction to the HC data with participants showing a massive increase of 0.38 for the Competitor (LC) target (mean: 0.55) compared to the Predicted (HC) target (mean: 0.17) ($F1 (1,28)= 154.7$, $MSE= 5.49$, $p<0.001$; $F2 (1,27)= 96.19$, $MSE= 5.5$, $p<0.001$).

There was a notable difference observed between the fluency conditions for the Competitor (LC) target: Participants had a 0.1 increase in proportion of fixations on the LC picture when hearing the Competitor (LC) target following a disfluent context (mean= 0.60) compared to a fluent context (mean=0.50). The proportion of fixations on the HC picture were roughly equivalent, with only a 0.02 difference after hearing a Predicted (HC) target following a fluent (mean= 0.18) and disfluent (mean= 0.16) context. This very small difference is in the opposite direction to the larger difference seen for the Competitor (LC) target. This resulted in a marginal interaction effect significant by items. This interaction reached significance in the by items analysis ($F1 (1, 28) = 3.92$, $MSE= 0.08$, $p=0.058$; $F2 (1,27)= 4.86$, $MSE= 0.09$, $p=0.036$). There were no effects observed for either by participants or by items analyses for the SR data, which is not graphed here (all $Fs <2.6$).

3.8 Discussion

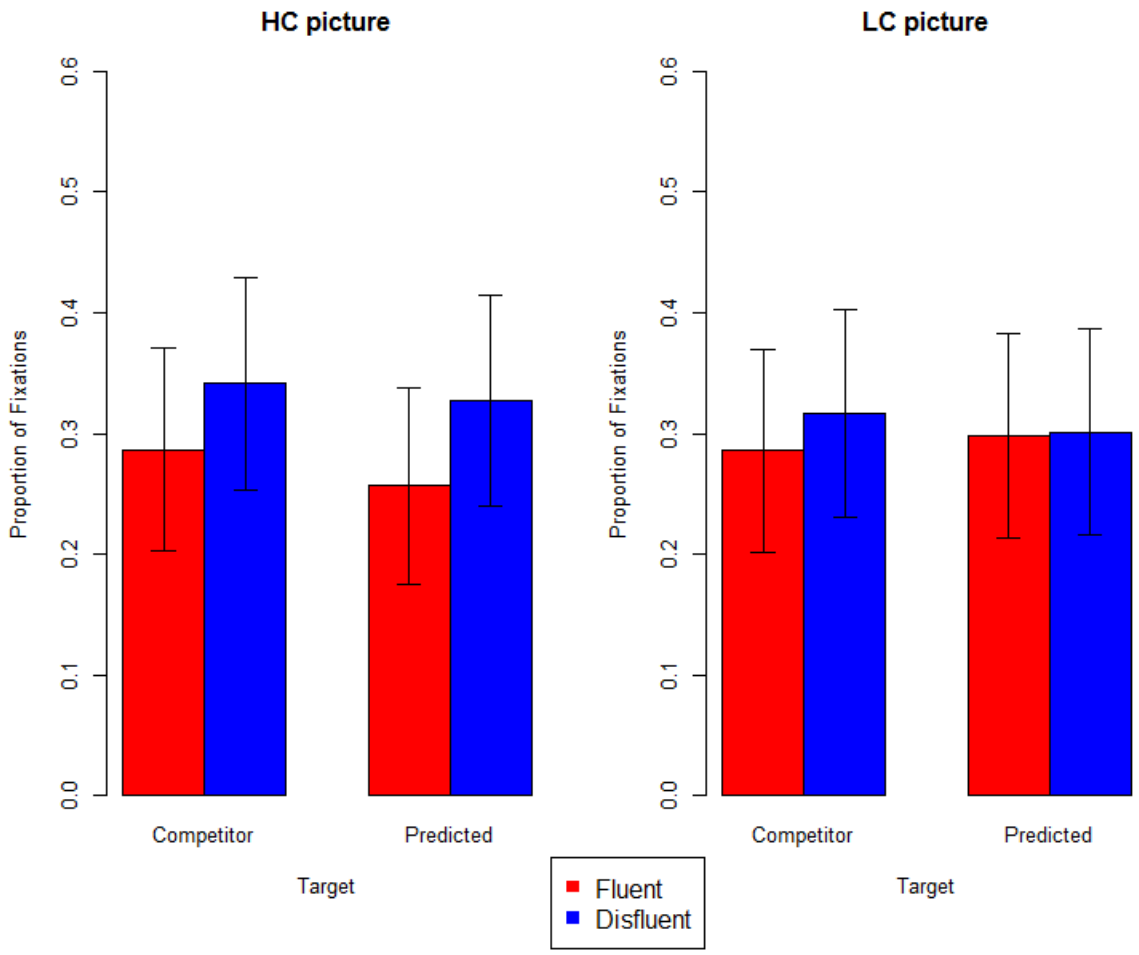
In the critical pre-TO period that our predictions were based on, following a disfluent sentence context there was an effect of participants making an increased proportions of fixations on the HC picture. This effect does not match up to either of the patterns of fixation behaviour predicted by a predictional or attentional account of disfluency processing during comprehension. The results for the current experiment also showed a lack of difference in fixation behaviour between fluency conditions for the LC picture. Both accounts would have again predicted an increased number of looks following disfluency. These results taken together counter and fail to replicate the many convincing disfluency effects previously seen during comprehension that support the predictional standpoint of disfluency

processing, where in the face of disfluency participants alter their online expectations resulting in looks to an unfamiliar referent within a scene (e.g., Arnold et al., 2007).

In the current experiment the only other plausible referent within the presented scene was the LC item and no effects occurred for this referent. The pattern of results seen for the SR picture do not support an attentional account based processing effect as there was an increased number of looks to this referent following fluent contexts. This result is in the opposite direction to the predictions made by the attentional account, which would anticipate an increased number of looks to the SR referent following a disfluent sentence context.

The pattern of results seen for this eye-tracking study does not provide any evidence to speak to the core question of differentiating between the underlying mechanism of disfluency processing provided by the predictional and attentional accounts. Furthermore, the results observed for the current experiment are hard to reconcile with the established empirical evidence on disfluency processing afforded by both accounts. This leads us to consider what influenced the experiment to create the pattern of results recorded for the current study. What is driving participants to increase the proportion of fixations towards the HC picture?

Having shown during pre-testing that the HC referent is the most predicted ending for the sentence contexts it is unlikely that the increased number of looks towards the HC picture following disfluency can be attributed to a lack of predictability in the sentence contexts.


HC picture
LC picture
Proportion of Fixations
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Competitor
Predicted
Target
Fluent
Disfluent

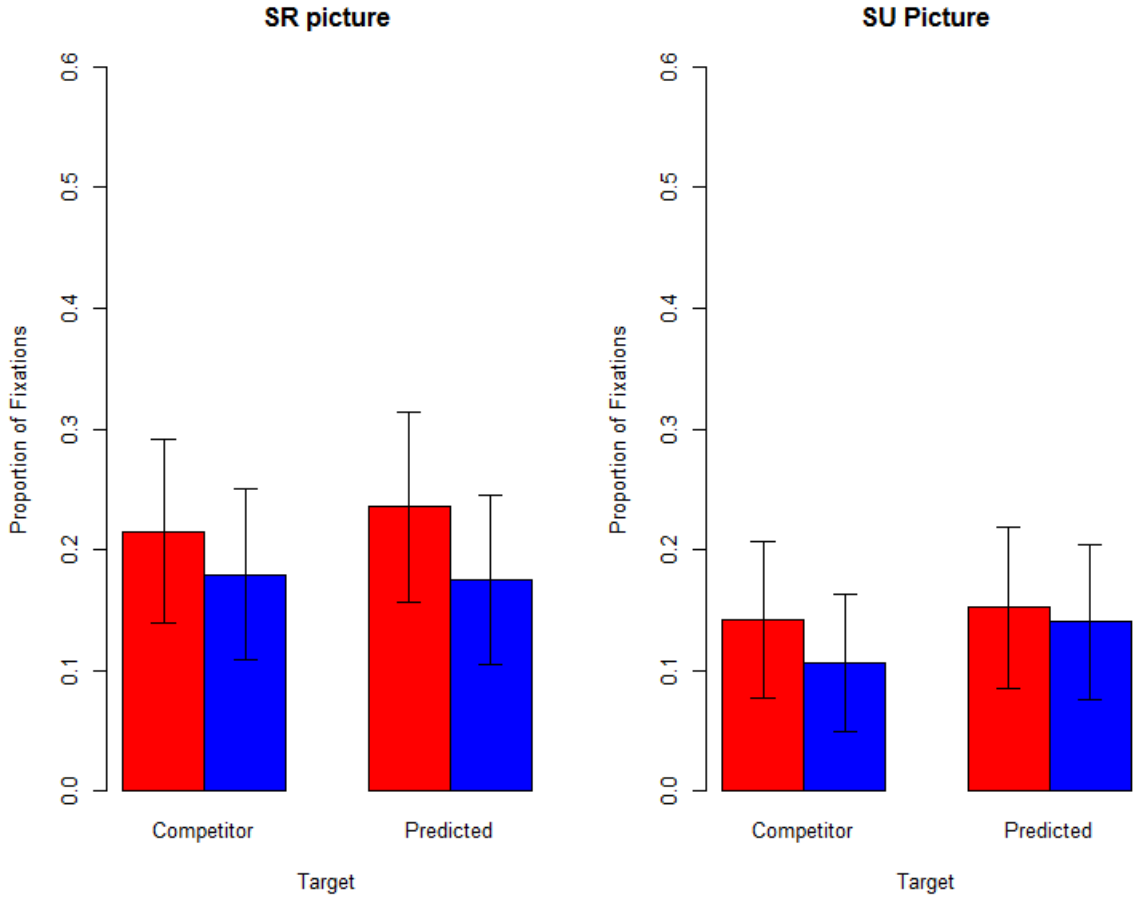

SR picture
SU Picture
Proportion of Fixations
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Competitor
Predicted
Target

Figure 3.6 (previous page)-The fixation percentages for all objects in a visual scene by role during the time period from 1000ms before until Target Onset (Pre-TO) by fluency of sentential context (Fluent vs. Disfluent) and by Target (Predicted vs Competitor).

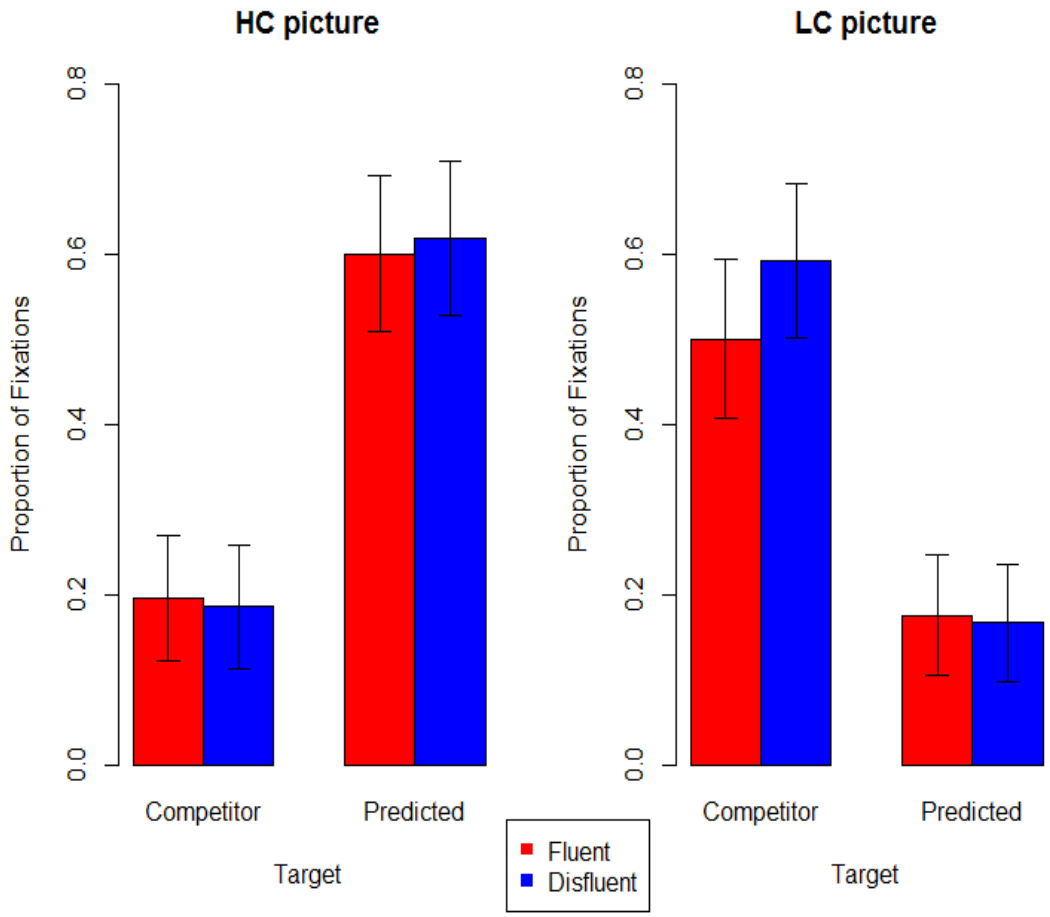


Figure 3.7 -The fixation percentages for all objects in a visual scene by role during the time period from 600-1000ms after Target Onset (Post-TO) by fluency of sentential context (Fluent vs. Disfluent) and by Target (Predicted vs Competitor).

We would expect that in the fluent conditions as the most predictable ending the HC picture would be most likely to receive the highest proportion of fixations but for the current study the average proportions were actually marginally higher for the LC picture following a fluent sentence context.

A simple reason for the results observed would be that participants are not sensitive enough to the difference between images to categorise them as competing entities. However for the second, post-TO time period analysed, the results followed the

predicted pattern showing that participants can distinguish between pictures. During this time period the target referent was heard and if they struggled to match the target heard with a referent then we would expect a spread of fixations but this is not the results seen. The HC picture received a highly robust amount of extra fixations in this post-TO time period when the Predicted target was heard following both a fluent and disfluent context. This was expected as the HC picture matches the predicted target of the sentence context. The opposite pattern of results was seen for the LC picture for the same post-TO period with participants showing a clear increase in fixations upon the LC referent following a Competitor target being heard in both a fluent and disfluent context. Again this was expected as the LC referent matched the Competitor target. For both the HC and LC referent participants were sensitive to the target heard and their patterns of fixations matched this. This replicates a well-established effect within the visual world literature for isolated pictures that when a referent is heard participants fixate on it (e.g., Altmann & Kamide, 1999). This sensitivity to the target distinction rules out that the participants were simply not sensitive to images and could clearly differentiate between at least the predicted and competitor images.

The current data leave further explanations for the pattern of unexpected results observed which are explored below with post-hoc tests that investigate a number of factors which may have influenced the outcome of the results seen. The first explanation focused on below is that participants were not sensitive, or showed limited sensitivity, to the disfluent token used in the experimental materials. If there was a lack of or limited sensitivity, then the differences between fluent and disfluent contexts would be minimal and the patterns of looking to each picture (HC, LC and SR) seen in the main experiment could be driven by other factors. A second linked explanation investigated below is that the disfluency effects may have been hidden by a lack of sensitivity from the paradigm and that the disfluency effect seen for the HC picture is erroneous. The results of the current study are discussed in relation to the findings of the post-tests in the general discussion below.

3.9 Post-Hoc Tests: Experiment 1.2- Audio cloze

In the main experiment we were interested in participants' eye-movement behaviour following a disfluent sentence context and how disfluency would affect the looks to each role of picture (HC, LC and SR) for this study but it produced some unexpected results. It was important to establish a reason for the lack of disfluency effects seen for the LC picture and the unexpected effect seen for the HC referent. One possible reason focused on here, was that participants were not sensitive, or showed limited sensitivity, to the disfluency used in the experimental materials. If this was the case, then a lack of difference between fluent and disfluent contexts in the LC picture would be expected and the patterns of looking to each picture (HC, LC and SR) seen in the main experiment could be caused by other factors, as there would be little separating the two fluency conditions. To gauge sensitivity to the disfluency, we tested whether participants would react to the disfluency used in the main experiment in a different paradigm.

An audio cloze test was chosen as there was a clear link to the pre-testing paradigm that used a written cloze test (e.g., Taylor, 1953) as the basis for the strength of context between a referent and the carrier sentence used in the main experiment. The outcome from this cloze pre-test resulted in the choice of referent to fill the HC and LC referent roles and had direct implications for predictions made for the main experiments about how disfluency would theoretically influence a participant's looking behaviour to HC, LC and SR referents within a scene. However, the difference in the modality of responding between the main experiment and the audio cloze task could highlight the effect of or sensitivity to disfluency in a different context. The audio cloze paradigm links participants' sensitivity to disfluency on measures of speech latency and spoken response categorisation which was different to the eye-movement measures employed in the main experiment. Within this paradigm, sensitivity to disfluency would be demonstrated by

participants producing different completions or a difference in speech latency between the fluent and disfluent conditions. Differences in either of these measures would highlight that the inclusion of disfluency in a sentence context stimuli had led to a change in behaviour.

This paradigm could not differentiate between our two post-disfluency predictions and their relative accounts as there would be no divergent behaviour for either account, for either of the testable measures. For token production, if participants produced a different token following a disfluency then it could be because of a disruption to their predictive processes or due to increased attentional resources or a combination of the two. However, it is not presently clear why differences in either prediction or attention based processing would lead them to change their own token, as at all times they know that they themselves are producing the token. Although they could model the speaker as having difficulties and they may not expect an upcoming disfluency, this need not affect the token they produce, so any effect could not be attributed to either predictive or attention based processes. Although this task cannot differentiate between the accounts being investigated, it can test if the results seen above were due to the task or the materials.

The audio cloze test ran as follows: Participants heard a sentential context, which was missing a sentence final object which they had to complete with their own production when given an onscreen prompt. It followed the design of the traditional written cloze test, as described above in the norming studies, but with an audio presentation. Using the example used for the main experiment above, participants would hear the "*The vet was sad to do so but he had to put down the local family's trusty…*" and they would then be required to produce a word to complete the sentence.

3.9.1 Participants

A total of 16 students from the psychology community at the University of Edinburgh psychology participated for course credit. Participants self-reported that they were native speakers of English and had normal or corrected to normal vision.

Students who had taken part in any of the pre-tests or the main experiment were excluded from participating. Two further participants' data was excluded, as they did not respond in a valid manner for over half of the trials and seemingly struggled to complete the required task.

3.9.2 Design and Materials

The same audio stimuli used in the visual-world experiment were used, as detailed above. The design of the audio cloze test was based on the 48 sentential context items from the visual world experiment. The 28 experimental items were used to create 2 lists: each list contained all 28 items but the presentation was either fluent or disfluent and varied between lists. Meaning both lists had a balanced number of (14) fluent and disfluent items. The 20 fillers from the visual world experiment were also included, these did not vary between lists and again half (10) were of each fluency condition. Fillers were included so that the pattern of the location of the disfluency in the experimental trials was made less obvious by the fillers. The disfluency always occurred immediately prior to the sentence final production for the experimental trials, whereas, the disfluency appeared in a number of pre-nominal locations throughout the sentential contexts in filler trials. Participants only completed one of the lists. Participants answered 4 practice filler trials before the main experiment and these were identical for all participants.

3.9.3 Apparatus and Procedure

The visual and auditory stimuli were presented using DMDX (Version 4.0.6.0, Forster, 2012) using a laptop PC with a 1366x768 pixel screen. The headset used was a logitech headset, commonly used for video calls.

Participants had to read an information sheet and then fill in a consent form. They were then seated in front of a laptop computer and had to place on a headset, which contained earphones and an attached microphone. Participants read the experimental instructions and performed 4 practice trials, which followed the same structure as the real experimental trials, so participants could familiarise themselves

with the experimental procedure. The instructions stated that participants would hear some spontaneous speech which lacked a word at the end of the sentences. They were told to complete the sentence with a spoken word in a natural manner. Quick responding at a normal volume was stressed. Trials started with a count down marker of "###", "##", "#"after which the trial would begin. This countdown marker was used so that participants' attention would be cued to focus on the auditory stimuli from the beginning. If content was missed it could create issues when asked to complete the sentence. Combined with the onset of the auditory stimulus, "+++++" was displayed on the screen for the duration of the sentence context. At the offset of the sentential context, the visual cue changed to "******" prompting participants to begin their productions. On top of the auditory stimuli, the addition of this visual cue gave a secondary signal to the participants to produce their token. Participants then had 2500ms to answer before a trial timed out. After the trial, the participant was asked to "Press the SPACE BAR to move to the next item." After the practice trials, the participants were always given a chance to ask questions and the experimenter checked that they felt familiar with the structure of each trial. They saw the instructions again reiterating the task before moving on to the data collection phase of the experiment. The trials followed an identical structure to the practice trials described above. The participants then completed 48 trials, with an optional break halfway through the experiment. The study took approximately 20 minutes.

3.9.4 Analyses

This audio cloze task was designed to investigate participants' sensitivity to the disfluencies used in the main experiment. Filler items were excluded from the current analyses. The design of the audio cloze allowed 2 measures to be collected and each was independently analysed. The first, focused on the content of raw responses produced. We were interested in whether participants' lexical responses varied following a fluent or disfluent sentential context. Responses were transcribed by the experimenter. The responses were categorised as matching if a participant's

production was the same word (in either singular or plural form) seen at least once in both fluency conditions. Responses that could not be transcribed accurately enough to make a valid semantic judgement on the collected token were excluded and accounted for 6.9% of the data. Responses were categorised as 'No response' if the participant did not produce any audible attempt to complete the sentence; this made up 12.3% of the data collapsed across conditions.

The second measure was onset latencies; here we were concerned whether there was a difference between the word onset times of responses following a fluent or disfluent context. The list factor discussed in the analyses relates to the 2 versions of the experimental items that a participant could have heard. Onset latencies were measured from the start of the critical period, which began on the offset of the sentential context, to the initial onset of a participant's response. Praat (Boersma & Weenink, 2013) was used to manually calculate the onset latency of each trial. The latency time was measured from the onset of a participant's production by viewing the waveform for each response. This bypassed issues of false triggers that could have affected any voice key use in this process. Additionally, for trials where participants did not respond (12.3%) or did not produce a semantically valid word (6.9%) the data was excluded. The predictors used for the onset latency analyses were: Fluency (Fluent/Disfluent) which was within participants and within items and List (1/2) which was between items and participants.

3.9.5 Results

We first present the results of the audio cloze task itself (ie. The word responses produced by participants) before reporting the analyses of the onset latencies of the responses produced. Each analysis was broken down by fluency condition of the sentential context (Fluent and Disfluent) and the List a participant saw (1 or 2).

3.9.6 Results: Responses

The mean percentage of same responses collapsed across conditions and fluency as shown in Table 3.3 was high (65.2%). The similarity of lexical responses collapsed across items was comparable following a fluent or disfluent sentence (t= -1.42, df= 54, p =0.16). However, the List condition an item came from did affect the similarity of lexical responses ((t= 3.05, df= 52.6, p<0.005). There was an increase of 15% in same responses for items from List 1 (53.9%) compared to List 2 (38.9%).

List	Same Responses (%)	Contribution by List (%)	No-Responses (%)	Contribution by List (%)
	65.2		12.3	
1	-	53.9	-	61.7
2	-	38.9	-	24.0

Table 3.3- Percentage of Same Responses, No-Response and the contribution of each condition to overall Same and No-response figure.

This variation is driven by the larger influence List 1 has on the number of same responses. As a lower number of participants took List 2 (7), compared to List 1 (9), there was some variation between responses expected. There is substantial by item differences in the production of same tokens, ranging from 0-89% of shared responses collapsed across List. The analysis of the response data here does not show any pattern of differences in the response data between fluent or disfluent condition. There is an effect for condition but this does not relate to any of the predictions made.

3.9.7 Results: Onset Latencies (Reaction Times)

As shown in Figure 3.8, disfluency reliably sped up participant's onset latencies in producing a response. A two-way ANOVA with Fluency and List as factors confirmed a main effect for fluency; participants onset latencies were on average 335ms quicker following a disfluency, [F1 (1, 14)=121.9, p >0.001; F1 (1, 27)=83.8, p >0.001]. There was no main effect for List (Fs <1). However, there was an interaction

effect between List and Fluency but only for the by participant analysis [F1 (1, 13)=14.2, p=0.002; F2 (1, 26)=2.3, p=0.14]. This interaction effect was driven by the disparity in size of difference between Fluency conditions within each List. The disfluent sentence stem latency average (542ms) showed a reduction of 260ms compared to the onset latency of the fluent condition (803ms) in List 1. In contrast, in List 2, there was an increased difference of 402ms between the fluent (875ms) and the disfluent (473ms) sentence contexts. The decreased number of participants in List 2 means there was less participants to average across. Therefore, each participant had a higher proportion of effect on the average, so this interaction effect is likely caused by greater participant variation across a smaller number of participants in List 2.

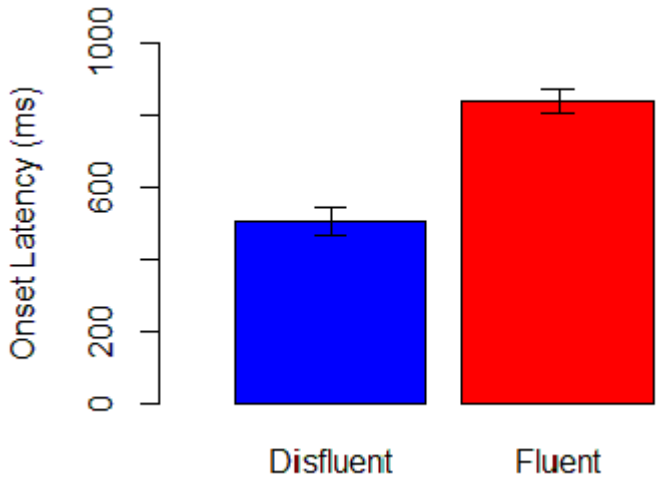


Figure 3.8 -By Participant means for onset latency (ms) following Fluent & Disfluent contexts.

3.10 Post Hoc Tests: Experiment 1.3 - Forced-Choice

As noted above, the pattern of results observed for the main experiment did not match our predictions. We suggested two reasons the pattern of results shown could be realised that we would test for in the post-hoc tests. Firstly, this pattern could have been driven by a lack of sensitivity to the disfluency used in that

paradigm, meaning that there would limited expected differences between fluent and disfluent conditions. The audio cloze post-test above tested participants' sensitivity to the disfluency used in the main experiment in a different task.

A second related reason, explored here, was that disfluency effects may have been too small for the eye-tracking paradigm to pick up. The second post-test again used a different task with the same experimental scenes to investigate participant's sensitivity to the disfluencies and address issues that may have arisen due to the pictures used in the experimental scenes. A forced-choice paradigm was selected. This paradigm employed identical materials and a similar procedure to that used in the visual-world experiment. The main difference was that participants had to press a button relating to the picture named in the sentence as quickly as possible instead of clicking on the correct referent. Participants were not eye-tracked during this experiment and the dependent variable was the response latency of the button press.

A forced-choice button press paradigm was chosen as there is a strong link to the main experiment; the same stimuli but with a different task. This means that sensitivity to disfluency could be tested using a different measure that is unrelated to eye-tracking. If participants' were sensitive to disfluencies presented in sentential contexts there would likely be differences in the response latencies following fluent and disfluent conditions. Firstly, the disfluency could help or hinder word recognition and hence, picture selection. Fox-Tree (2001) found that post disfluency listeners were quicker to identify a target word in a sentence context than in a version without a disfluency present. Corley and Hartsuiker (2011) found that following a filled pause participants were quicker to select a referent from two pictures with a button press, in comparison to a fluent equivalent. The current study employed 4 pictures but was similar in the goal action of selecting a visually presented referent from an auditory target. From the evidence provided from these studies we would expect disfluency to speed up participants' picture selection in comparison to the fluent conditions.

Secondly, we want to differentiate between the predictive and attentional accounts of disfluency processing. Both accounts were tested using the same disfluency for the current experiment, so would be equally likely to experience the speeding effect described above. However, neither Fox-Tree (2001) or Corley and Hartsuiker (2011) featured contexts that predicted one or more referents in the scene. Therefore, the interaction between disfluency and these predictive processes or lack thereof, could create different patterns of onset latency behaviour. In fluent trials, if participants were to employ predictive processes then we would expect them to be predicting the HC picture as the referent they would have to select. If the predicted target (HC) is heard then this would be likely create the quickest response latency. In this 'match' condition, the expectations that unfold as the sentence progresses match the realisation of target. Therefore participants would only require minimal auditory stimuli from the target to confirm their prediction. A 'mismatch' condition would occur in the fluent trials when the participants heard a competitor target (LC), as they are likely to have been biased towards predicting the HC picture. This prediction would only be violated upon heading the target onset of the competitor target (LC). Therefore, we would expect participants to take a longer time select a referent following a competitor target (LC).

After encountering a disfluency, if participants were to alter their predictive processes based on speaker modelling then we would expect them to realign their predictions to the only other plausible picture in the scene the LC picture. Therefore, when the Competitor target (LC) is heard post disfluency, this creates another 'match' condition. In both ‘match’ cases, the same predictive processes would have narrowed down to a unique target before participants had to select a picture, so you would expect similar latency times. However, if a Predicted target (HC) was heard post disfluency then this would create another 'mismatch' between participants’ expectations and the target heard. Therefore, we would expect participants to take longer to select a referent than in the 'matching' fluent baseline and a comparable time to the 'mismatch' fluent condition.

If participants are not making any predications and instead, employing more attentional resources after a disfluency, as stated by the attentional account, then we would predict speeded responses post disfluency, as in Corley and Hartsuiker (2011). Crucially, we would expect balanced response times across the targets heard as the lack of predictive processes makes either target (Predicted (HC) or Competitor (LC)) equally likely and there are no predictions to be matched or violated. A third option would be a lack of difference between fluent and disfluent sentential contexts. The results are discussed in relation to these predictions below. Overall, divergent behaviour between fluency conditions would signify an interesting effect.

3.10.1 Participants

A total of 17 students from the University of Edinburgh community participated for a monetary reward. Participants self-reported that they were native speakers of English and had normal or corrected to normal vision. Students who had taken part in any of the pre-tests or the main experiment were excluded from participating.

3.10.2 Design and Materials

The design of the forced choice paradigm was identical to the main visual-world experiment until the picture selection phase; in the main experiment participants had to click the picture heard using a mouse. However, for the current study participants had to select the picture (either HC or LC) that matches the target they heard by pressing a button on a keyboard. The design was 2 (fluent vs. disfluent) X 2 (Predicted or Competitor Target) and both variables were within subjects and items. Participants were randomly assigned to one of four conditions that matched the conditions detailed above in the main experiment.

3.10.3 Apparatus and Procedure

The visual and audio stimuli were presented using 'Experiment builder' software (version 1.10.165, SR Research, 2012) on a PC and a 15 inch monitor set at a 1366x768

resolution. The keyboard used was a standard Dell English Language keyboard. Participants then had to press any key to move to the next trial. There was an optional break half way through the experiment.

After reading an information sheet and filling in a consent form, participants were seated at a computer screen. Participants then read through the instructions presented onscreen. The instructions included a brief description of the task and stressed quick and accurate responding. At this point the experimenter instructed the participants to adopt a certain hand position on the keyboard, which placed fingers on the 4 keys that could be used to select a picture. They were also told to keep this hand position for the duration of the experiment. This was to give equal access to all keys throughout the experiment. The keys used for responding were located in the number keypad ('1', '3', '7', '9') and the location of the key corresponded to the picture in the same spatial location; for example, the '1' represented the lower left quarter of the screen.

Following this they performed 4 practice trials, which did not vary across participants. These practice trials comprised of 4 filler trials taken from the main experiment and were designed as a familiarisation phase. The trial began with the experimental scene being presented for a second before the sentential context began playing. Immediately after the sentential context finished the target was played. Upon completion of the target, an orange circle appeared in the middle of the screen. At this point the buttons became active and participants had to press a button to select the picture corresponding to the target heard. The practice trials followed the exact same structure as the main experiment. After the practice trials, the participants were always given a chance to ask questions and the experimenter checked that they felt familiar with the task. After a repeat of the set of instructions issued at the beginning of the practice trials, participants began the data collection phase of the experiment. They were also informed that there would be an optional break during the experiment. Participants then completed 48 trials. The study took approximately 30 minutes to complete.

3.10.4 Analysis

The focus of the current forced choice task was to investigate the influence of disfluency on the time taken to react and push a button selecting one of the 4 pictures after a target word is heard. The measures used were the time it took participants to press a button after the offset of the audio target and the selection of the correct referent. The reaction time was generated from a report produced automatically by the eye-link data-viewer software. Trials where participants had selected the wrong referent were removed and this accounted for 6.8% of the data. The predictors we use in the analyses are fluency condition (Fluent/Disfluent) and Target (Predicted (HC)/ Competitor (LC)) which were within subjects and items. Additionally, the Condition the participant was assigned to (1-4) was used as a factor which was between participants and items.

3.10.5 Results

There was little variance between fluent and disfluent conditions: Disfluent presentations were on average 23ms quicker than fluent presentations. Figure 3.9 shows the average reaction time by Fluency Condition. A three-way ANOVA with Fluency, List and Target as predictors, showed no main effects (all Fs <2). However, there were a number of interactions: There were notable differences by fluency across conditions [F1 (3, 13) =4.44, p= 0.02; F2 (3, 98) =4.84 p= 0.004] as seen in Figure 3.10.

A second significant interaction was observed between target and Condition [F1 (3, 26) =5.87, p= 0.004; F2 (3, 98) =3.49, p= 0.02] as seen in Figure 3.11. Neither of these interaction effects speaks to our predictions, so are not reported in any more detail.

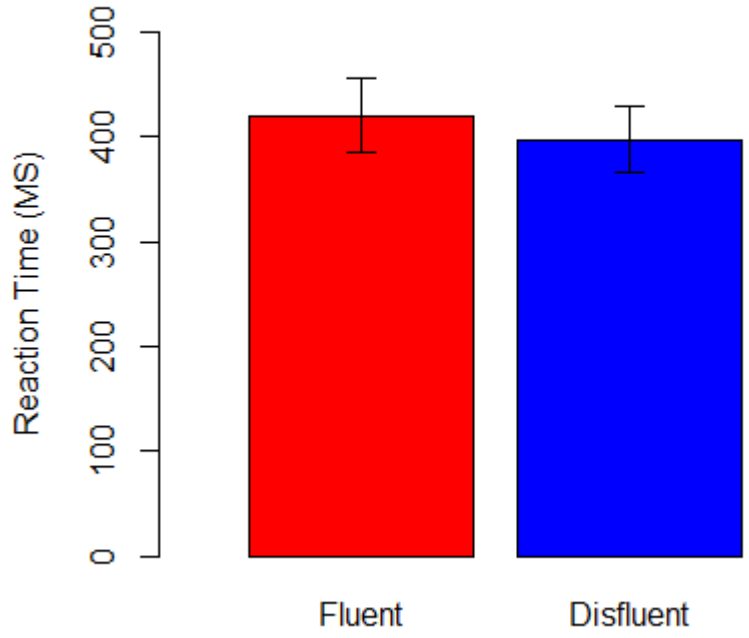


Figure 3.9- By participant means for reaction times (ms) following fluent and disfluent contexts.

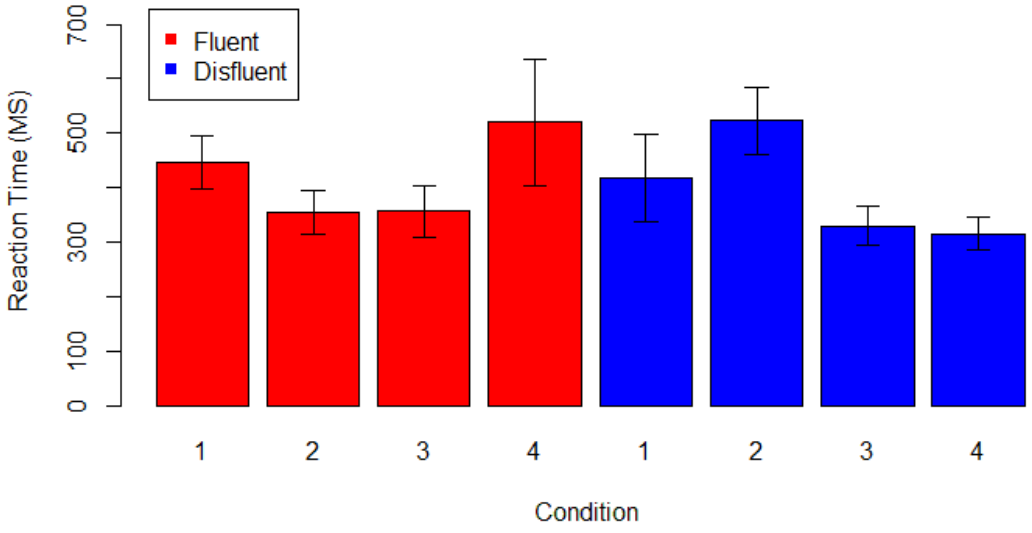


Figure 3.10- The average reaction time (ms) by Condition and Fluency word.

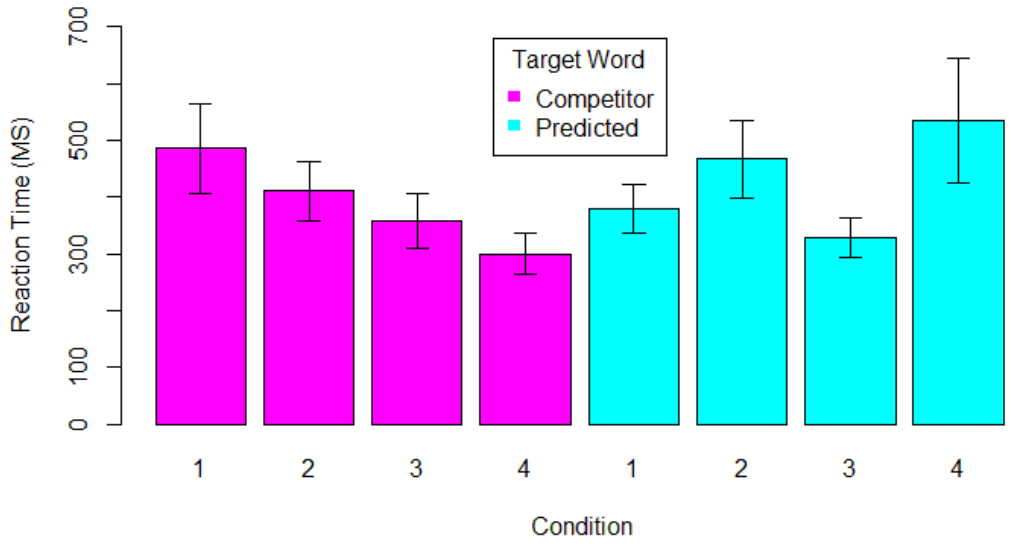


Figure 3.11- The average reaction time (ms) by Condition and Target word (Competitor (LC) & Predicted (HC)).

3.11 General Discussion

In the main eye-tracking experiment we found an unexpected disfluency effect that did not match the expected pattern of fixations predicted by either the predictional or attentional accounts of disfluency processing, as detailed in the discussion of Experiment 1 above. Instead we found that following a disfluent utterance there was an effect of participants making an increased proportions of fixations on the predicted target (HC) picture during our critical pre-target onset (Pre-TO) time period. Additionally, there was no reliable variance between fluency conditions for the proportion of fixations to the competitor target (LC) picture. We ran two post-hoc tests to help explore the results seen for the main experiment. These post-hoc studies investigated further whether participants were sensitive to the disfluency in the paradigm and whether the paradigm could adequately capture any disfluency effects.

The first of the post-hoc tests was an audio cloze test. This novel test was employed to test participants' reaction to the materials used in the main experiment but in a different task. Participants were asked to produce a word to complete a fluent or disfluent version of the utterances used in the main experiment. The results for the raw responses for this test showed that the words being produced did not vary as a function of whether they were preceded by a fluent or disfluent utterance. However, as stated in the introduction to this study, it was not clear that even if participants were sensitive to the disfluency whether this would lead participants to change their production, as at all times they knew that they were in control of producing the token. Although the instance of disfluency could lead them to model the speaker as having difficulties, it may have also been the case that disfluency was redundant in this task of producing a response and need not affect their production or be reflected in the token they produce.

The second measure, onset latency, revealed a reliable speeding up of participants' production following disfluency; participants were 335ms quicker. This significant difference in onset latency times suggests that participants were sensitive to the disfluency being heard in the main experiment. However, the speeding up could have also been driven by the extended processing time afforded by the disfluency allowing participants extra formulation time leading to a quicker response following the offset of the disfluency. A temporal delay has been seen to speed up word identification (Corley & Hartsuiker, 2011) but the effects seen in the Corley and Hartsuiker study were not for trials where the participants had make their own production and had fixed outcomes for each trial. In relation to the predictional and attentional accounts of disfluency processing, the audio cloze task could not differentiate between accounts as there would be no divergent behaviour to separate either account, for either of the testable measures.

The second of the post-hoc tests was a forced-choice button press study. This paradigm had a strong link to the main experiment: the same stimuli but with a different task. This means that sensitivity to disfluency could be tested using a

different measure that is unrelated to eye-tracking. The results showed there was no reliable difference in time taken to select a referent following a fluent or disfluent utterance. There was a 23ms advantage for the disfluent utterances but this was not even marginally significant. It is possible that with increased power this may have reached significance, as the difference is not too far away in comparison to the magnitude of effect found in Corley and Hartsuiker (2011), where a difference of 32ms was significant. However, in their study they controlled for temporal length and included disfluency in the control condition.

Taken together these post-hoc tests provide mixed evidence about participants' sensitivity to the disfluency in the main eye-tracking experiment. For the speeded production observed in the audio cloze test the lack of duration matching between fluent and disfluent utterances creates a confound which is hard to account for. It may be the case that the disfluency present within an utterance is driving the decreased onset latency but from this paradigm it cannot be differentiated from the extra processing time that it allows participants.

Neither the predictional or attentional accounts of disfluency can be reconciled with the pattern of results seen in the current eye-tracking study. A possible explanation for the unexpected increase to the predicted target (HC) following a disfluency for the main experiment is that disfluency processing is flexible and dependant on the task being undertaken. We propose a combined predictional and attentional account that we will call the Combined account. The attentional account proposes that following a disfluency predictional processes are stopped and the additional attentional resources are used to focus on bottom-up processing to facilitate comprehension. However, the results from the main experiment suggest that participants are continuing to predict the upcoming referent based on the contextual fit of the preceding utterance. The Combined account of the attentional account proposes that the heightened attention seen following a disfluency (e.g., Collard, Corley, MacGregor, & Donaldson, 2008) can be complementary to other processing such as predictional effects (e.g., Arnold et al., 2007; Heller et al., 2014).

This Combined account could work in one of two ways: The first is that increased attentional resources can be used to facilitate listeners to attend to either top-down or bottom-up processing based on the situational need created by the context of the utterance or the task to maximise the chance for successful comprehension. Variable attending to the incoming speech signal has support in the speech perception literature, which has shown that when subjected to an increase in cognitive load listeners demonstrate an increased reliance on lexical influences (Mattys & Wiget, 2011). However, when focused attention is directed to the speech stream, as during a lexical decision task on ambiguous stimuli, listeners show an increased sensitivity to the bottom-up processing of fine acoustic detail (Pitt & Szostak, 2012). A second linked possibility is that following disfluency or an uncertain delivery, listeners similarly use the increased attentional resources to attend to the bottom-up signal but they are actually *more* sensitive to the top-down predictional effects. So instead of being a trade-off being bottom-up and top-down processing, it is actually an automatic ramping up of the whole perception system to maximise the chance of comprehension when the speaker seems unsure of themselves. The core difference between these two proposals is whether following a disfluency, the listener can modulate the control of attending to top-down processing in unison with increased bottom-up processing.

Both possibilities could account for the increased proportion of fixations seen for the predicted target (HC) in the eye-tracking paradigm. Following a disfluency, users are making increased use of the bottom-up signal and the preceding sentence context that is heavily biased towards the predicted target (HC) which explains the increased number of looks towards this picture. Additionally, as noted in the norming study, our HC and LC items were not rated as matched in plausibility and this could have influenced participants' strength of prediction towards to the HC item. This could also explain the lack of any fluency effect for the competitor target (LC) picture, as with the increased reliance on predictional processing in the

disfluent condition; we would not predict any difference in fixation proportions between fluency conditions.

Our findings were unexpected and did not match up with our predictions or either of the disfluency processing accounts that we aimed to investigate. Instead, we propose a Combined account that suggests that disfluency processing may be better explained with the possibility of variable attending or automatic increased reliance on both bottom-up and top-down processing to maximise successful comprehension.

CHAPTER 4

Experiment 2

4.1 Chapter Overview

In the previous chapter, the results of the eye-tracking study were not predicted. They did not differentiate between the predictional and attentional accounts of disfluency processing. In the current chapter, we further investigate the attentional account of disfluency processing during language comprehension using a speech perception paradigm.

The attentional account has support from a number of empirical findings, as detailed in the literature review above and outlined here. Fox Tree (2001) showed that following *uh*, participants were quicker to identify a target word than in the related condition that featured a silent pause of the same duration. Fox Tree proposes that this facilitation effect is driven by a heightening of attentional resources following the disfluency. However, participants did not take less time to respond following *um* than following a silent pause. Fox-Tree suggested that *um* represented an increased delay in upcoming speech, meaning that orienting attention would be impractical when the time course of the resumption of speech is unknown. Corley and Stewart (2008) proposed an alternative explanation, they observed the duration of the remaining silent pause from the excised *um* represents a break in speech that extends beyond a normal gap found for a fluent delivery, therefore, the silent pause in the *um* condition may also have been comprehended as disfluent or processed in a manner divergent from typical fluent speech.

Collard, Corley, MacGregor, & Donaldson's (2008) showed that there was a notable decrease for the attention based P300 brain component when a novel stimuli was presented following a disfluency.
The reduction seen for this attentional component suggests that participants were already attending to the incoming speech; providing support for the viewpoint that

following a disfluency, listeners are orienting their attention to the upcoming content and this heightened attention is responsible for facilitation effects seen following filled pauses (e.g., Brennan & Schober, 2001; Fox Tree, 2001). Taken together these studies provide a persuasive body of work that support an attentional mechanism being heightened during disfluency processing in comprehension. The attentional account suggests that these facilitation effects following a disfluency are causing listeners to abandon predictional processes and rely on the incoming speech signal, bottom-up information, to resolve the comprehension difficulty posed by the interruption to the speech.

Within speech perception attention has been shown to impact upon a listener's perceptual sensitivity to fine grained acoustic information (Pitt & Szostak, 2012; Cutler et al., 1987). Taken together these studies showed that increased attention to incoming speech resulted in an increased likelihood for listeners to rely on the bottom-up acoustic signal heard. These effects are discussed in more detail in the literature review above and in the introduction below. In the current studies we explore whether following a disfluency there is heightened attention by employing a speech perception paradigm. If disfluency were to drive increased attention to the incoming speech stream then this would be expected to impact upon low-level speech perception resulting in a similar increased sensitivity to incoming speech stimuli, as shown for the attention based perceptual effects. The results of this study are discussed in relation to the implications this has for the attentional account of disfluency processing.

4.2 Introduction

As outlined above, there is evidence to support an attentional account of disfluency processing that states that attention is heightened following a disfluency (e.g., Collard et al., 2008; Fox Tree, 2001). Attention has also been proven to be influential

during speech perception processes. This topic is detailed further in the literature review.

In an experiment which required listeners to make lexical decision about single word targets, Pitt and Szostak (2012) found that by varying the attentional focus of listeners, through the instructions read at the beginning of the task, they could induce changes in the proportions of ambiguous sounding stimuli that were categorised as 'words'. The explicit attentional manipulation employed was the task instruction that participants saw; "Participants given the focused instructions were informed that the ''s'' or ''sh'' letter sound in a particular word position could be ambiguous, and that they should listen closely so as to make the correct response" (Pitt & Szostak, 2012: 1229). The unfocused instructions did not signpost the target phoneme or location within a target word or not. Participants had to make lexical decisions on words that had the ambiguous phoneme placed in initial, medial and final locations within words. A lexical bias was seen in the unfocused condition across the range of phoneme locations within a word, with participants showing a tendency to label ambiguous stimuli as words. However, in the focused attention condition participants labelled a lower proportion of target stimuli as 'words'. The findings here confirm an attentional effect and provide additional evidence that participants can exert control of how they attend to contextual and fine grained acoustic information during perceptual processing. Pitt and Szostak align with Mirman et al.'s (2008) proposal that attention acts to damp lexical influences, further supported by the increasing effects of attention across word positions. Taken together, these studies provide clear evidence for the influence of an attentional modulation in speech perception using either task demands or an attentional manipulation.

The attentional manipulation noted in the Pitt & Szostak (2012) study provides a testable prediction for the attentional account to be measured against. We propose that if disfluency affects attention then it may impact processing in the same way as

the explicit instructions employed in the Pitt and Szostak paradigm. If there is heightened attention post disfluency then this should drive increased attending to the incoming fine-grained acoustic detail in the speech stream resulting in a pattern of findings that match the focused instruction condition in the Pitt and Szostak paradigm. We would predict if the attentional account is correct then there should be a similar reduction in the proportion of 'word' responses seen post disfluency in comparison to the focused instruction condition.

Although, the Pitt and Szostak paradigm will form the basis for the current study, a disfluency occurring before a single word is not representative of how disfluency typically occurs in everyday speech and could have impacted upon participants' processing during the lexical decision task. We therefore created neutral sentence stems to precede the target words that contained the word-initial pronunciation variation. This created a pre-target juncture for the inclusion of disfluency that would be more reflective of disfluency use in everyday speech. The variation in the speech took place along the same frication continuum (/s/-/ʃ/) employed in Pitt and Szostak (2012).

The current study was the first in a series of experiments. Our primary concern in this first study was being able to show that we could replicate the attentional effect recorded in Pitt and Szostak (2012) with our paradigm and materials before the inclusion of disfluency. A lack of effect with the inclusion of disfluency would create an attribution problem, as the result may have been driven by the presence of disfluency or our untested sentence based paradigm and materials. Therefore, the current study employed the method of instruction at the beginning of the study as the primary attentional manipulation with a fluent production, as previously used in Pitt and Szostak. It follows that based on the previous demonstrations of an attentional effect by Pitt and Szostak (2012) we would expect an increased reliance on bottom-up phonetic information for those participants who saw the 'focused' instructions. If this is the case, this would result in an increase in the proportion of 'non-word' responses in comparison to the 'unfocused' instructions. Hence,

attention would modulate their rating of lexical acceptability for a word. An additional lexical bias would be expected following the 'unfocused' instructions, as there would be nothing driving increased attention to the fine-grained acoustic information contained in the incoming speech stream.

4.2.1 Target Word Selection

The experimental targets were crucial for the current study as they carried the word initial phoneme variation that formed the basis of the predicted differences between instruction conditions, so we wanted to create a matched target pair. Our primary concern was the following phoneme sound being consistent across both words, so initially we opted for 'Sand' and 'Chandelier'. Pitt & Szostak used 'Chandelier' in their first experiment and we aimed to be follow this original study, as we would like to replicate their results here before adding in disfluency (2012). The matched word initial /s/ word Pitt used was 'Serenade' but we did not use this, as in British English the vowel following the /s/ phoneme is not realised as equivalent and we wanted to present the phoneme in the same context across the word pair. In our chosen words, the critical phoneme was followed by the same vowel, /æ/.

4.2.2 Continuum Creation

First, we obtained the /s/ and /ʃ/ phonemes that form the endpoints of the continuum by recording productions in a /ə/ context housed within words. The materials used to build the continuum were recorded at a University of Edinburgh studio facility by an engineer with the author present. All materials were produced by a native British English speaker. The speaker was instructed to produce the words containing the /s/ and /ʃ/ tokens in a natural manner. The recording of the current materials took place during the recording session that produced the rest of materials detailed below. All recordings were saved in a mono 48kHz .wav format. These /s/ and /ʃ/ phonemes were then isolated by excising this vowel sound. The /s/ and /ʃ/ phonemes were produced in a number of words, with the clearest example of each chosen to be used. The fricatives were then matched for duration and loudness. A 19 point/s/ to /ʃ/ continuum for the word-initial position was created by

digitally blending the two phonemes in varying proportions. This continuum creation method is adapted from that used in Pitt and Szostak (2012). All word initial fricatives were 147ms in duration after editing. This was shorter than the 215ms duration seen for the word initial fricative in Pitt and Szostak (2012). We aimed for a 5 step continuum with points 1, 3 and 5 set as fixed proportions: Point 1- 100% /s/; Point 3- 50% of each phoneme and Point 5- 100% /ʃ/. This left 16 intermediate steps in the continuum that were made by blending together the initial /s/ and /ʃ/ phonemes in different proportions in 5% intervals, creating 8 steps for the majority /s/ side of the continuum from which to select Point 2 and 8 steps for the majority /ʃ/ side of the continuum from which to select Point 4.

4.3 Norming Studies

The current study was designed to elicit theoretical differences in lexical responses along a word to non-word continuum using /s/ and /ʃ/ phonemes in a pair of target words: 'Sand' & 'Chandelier'. Therefore, we needed to make sure that the steps along our continuum both: (i) differed enough from one another to speak to our predictions and (ii) did not occupy the extremes of proportions: either 1 or 0 or values which were close to them, as this could mean responses would be subject to ceiling or floor effects.

4.4 Pre-Test 1

The first pre-test was concerned with testing each of the 16 intermediate continuum steps in the word-initial continuum to decide which of these would be used as points 2 and 4 in the final 5-step continuum. These intermediate points of the continuum needed to be spread across the range of proportions and still generate strong lexical bias.

A total of 11 students from the University of Edinburgh psychology community participated in the experiment. The 16 intermediate steps of the continuum described above were spliced onto the position-matched word parts: '_and' &

'_handelier'. These word parts were recorded during the same session as the remaining main experiment and continuum materials. The details are given below in experiment 2. The addition of the phoneme sounds created a continuum of target words each with the same 16 steps from the word-initial phoneme continuum. The duration of the 'Chandelier' targets was 802ms and the 'Sand' targets was 652ms.

Participants listened to the target words and had to provide a lexicality judgement on a written form by simply writing a tick or cross. The target words were presented in a random order. Participants were tested by the experimenter in a group environment after a tutorial. The participants completed the 32 trials whilst seated in silence and could not discuss their answers with their peers. The study lasted around 10 minutes.

4.4.1 Pre-Test results: Target variants chosen

We were interested in the strength of lexical bias for each continuum point and associated target word. The measure employed was the percentage of word responses for each variant of the target words. For the 'Sand' targets there was only one variant that demonstrated a strong lexical bias (over 70%) and this point was chosen as Point 2. For the 'Chandelier' targets there were 6 variants that elicited a very strong lexical bias (over 90%). These were the targets that included the highest proportions of the /ʃ/ phoneme. The continuum point chosen as Point 4 from this group demonstrated a high level of lexical bias while also providing a good amount of variation away from the nearest continuum points. This allowed the final continuum to be more evenly spread across the /s/ to /ʃ/ range. As a 90% lexical response rate was high, we predicted that focused attention would lower the number of word responses, so this figure would not be subject to a ceiling or floor effect.

4.5 Pre-Test 2

The second pre-test checked whether listeners could distinguish between targets that included different points along the word initial 5-step /s/ to /ʃ/ continuum that was created in the first pre-test. It was a variation on the well-established AXB method for testing for sensitivity between related phonemes (e.g., Boersma & Chladkova, 2013; Gerrits & Schouten, 2004).

A total of 10 students from the University of Edinburgh psychology community participated in the experiment. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties. Participants who had completed the previous Pre-Test were not allowed to take part. Participants were rewarded with course credit upon completion of the study.

Trials were made up of three repetitions of one of the target words, either 'Sand' or 'Chandelier', and the second (middle) word (the X part) could either be the same or different to those around it (the A part). So in each trial, the A and X parts could vary by continuum point.

During the experiment participants heard the full range of continuum points for both targets in both the A and X parts, so for each target there was continuum points 1-5 in the A position and for each of these A parts there was 5 corresponding continuum point 1-5 X parts, meaning that there were 50 trials: 25 'Sand' trials and 25 'Chandelier' trials. The trials were presented in a random order.

Participants were instructed about how each trial would sound and how they should respond, after which a test trial was played to them and they had the chance to ask any further questions before the main pre-test started. Participants were asked to judge whether the second repetition of the word was the same or different to the first and third repetitions and mark this on a sheet provided to them by the experimenter. Participants could ask for a trial to be repeated once, if for any reason they were not concentrating on the task or wanted to hear a trial again.

Participants sat at a desk in the lab alongside the experimenter who had a laptop computer which played aloud each trial. The volume of the laptop was kept constant for each participant and at this volume the audio could be heard clearly. The study took between 15-20 minutes to complete.

4.5.2 Results

The measures were the percentage of same responses for each trial and the continuum gap between the A part target and the X part target: If the A part target used Continuum point 1 and the X part target used Continuum point 4 then the Continuum Gap equated to 3 steps. There were not an equal number of observations for each Continuum Gap, with the number of trials for each gap decreasing as the gap increased. Figure 4.1 shows the percentage of same responses. This graph shows that for no-gap or for a gap of 1 place on the continuum participants were poor at distinguishing between these points. However, there was a sharp drop off from a gap of 2 steps along the continuum and above, showing that participants could effectively distinguish between these points

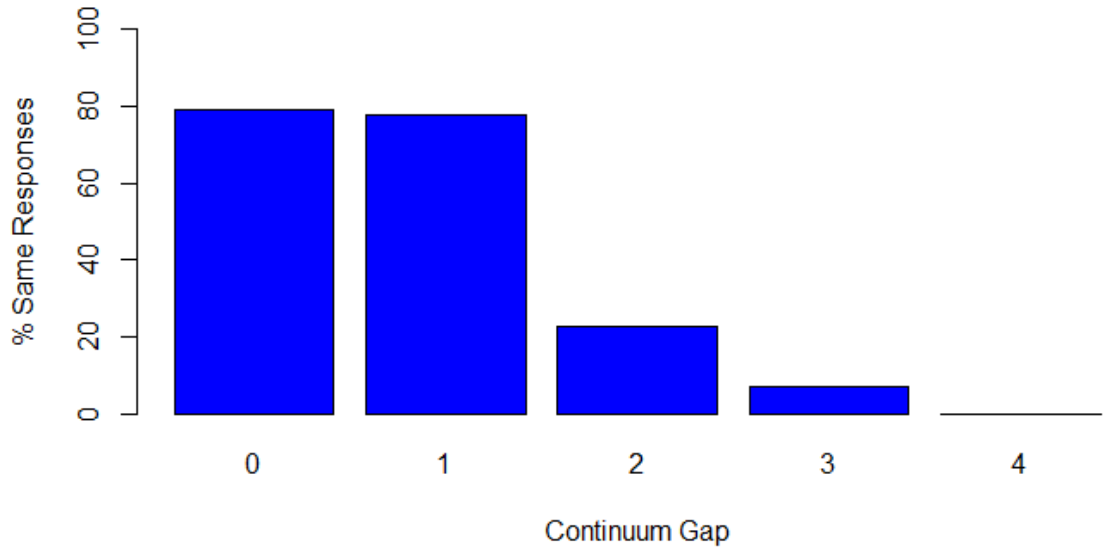


Figure 4.1- The percentage of same responses for each step of continuum gap for Pre-Test 2.

In the main experiment, participants did not have to judge targets against one another and the primary aim of this pre-test was to gauge participant's sensitivity to different places along the continuum. Clearly, participants could distinguish between places on the continuum that were not consecutive. This means that in the main experiment we can say with some certainty that each continuum point can be recognised as a unique entity that is distinct from other continuum points. Although, neighbouring continuum points are much harder to judge as being distinct. Conversely, if there were no differences found between continuum points in the main experiment then this is not down to participants perceiving all continuum points as the same.

4.6 Experiment 2: Speech Perception and Attention

In the current study, we ask how focused listener attention affects lexical decision at a phonemic level. Pitt and Szostak (2012) demonstrated that the effect of phoneme manipulation is reduced when participants' attention is explicitly directed to the ambiguous phoneme, with participants less likely to categorise an ambiguous item as a "word" under such conditions than otherwise. We applied this paradigm at the sentence level to investigate whether heightened attentional focus led to greater perceptual sensitivity at a phonemic level. Specifically, we compared the impact of attention on lexicality judgements when there was a phoneme manipulation.

4.6.1 Participants

A total of 26 students from the University of Edinburgh participated for a reward of £6.50 upon successful completion of the current and an additional short study. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties. Participants who had taken any of the pre-tests were excluded from taking part in the main experiment.

4.6.2 Design and Materials

Each trial was made up of a place holder sentence (e.g, '*I had lost the…'*) followed by a target word that participants then had to make a lexicality judgement on. There were two types of target word: experimental targets and filler targets. Following pre-testing we were concerned that the target words were not matched for length or syllable structure and this diverged from the Pitt & Szostak (2012) paradigm: 'Sand' was mono-syllabic and only 4 phonemes and 'Chandelier' was trisyllabic and 8 phonemes long. Therefore, we amended 'Sand' to 'Sandcastle' as we wanted both targets to be closely matched for these criteria. These changes meant that the target pair were both trisyllabic and matched in length and had word initial /s/ to /ʃ/ phonemes that were followed by the same vowel. Our revised /s/ target word, 'Sandcastle' contained the previous variant 'Sand' which had already been pretested; therefore, we did not rerun the norming process for the new variant. The duration of the 'Chandelier' targets were 802ms and the 'Sandcastle' targets were 981ms. The experimental targets were always derived from either: 'Sandcastle' or 'Chandelier'. The experimental targets each had 5 variants that corresponded to a change in their initial phoneme sound for each place of continuum described above. At one end of the continuum (as in normal productions) each experimental target was a word, whilst the same target was made into a non-word at the other end of the continuum: 'Sandcastle' became 'Shandcastle'; 'Chandelier' became 'Sandelier'.

There were 40 filler targets that formed 20 word pairs: 20 'word' fillers (e.g, holiday) and 20 non-word fillers (e.g, foliday). The 'word' fillers varied in their initial phonemes. Their matched non-word filler variants replaced the initial phoneme with another phoneme that only varied by either place or manner of articulation. Fillers varied in length from one to three syllables. The 10 place holder sentences were included to increase the ecological validity of the task. They were short and context neutral, so that participants were not anticipating any certain entity and so that they could work with both the experimental targets and filler targets. All

sentences were between 8-17 characters long and began with 'I' and preceded the target with the determiner 'the', for example, "I had seen the…". They followed the same structure so that the effect of the place holder sentence would be minimised.

Participants saw 10 blocks, with 30 items in each block, comprising 300 trials overall. Each block consisted of the 10 place holder sentences repeated three times. All 10 experimental targets were included in each block. 20 fillers were included in each block: 10 word fillers and 10 non-word fillers. All filler pairings were used in each block: a filler and its matched pair could not co-occur within block. A filler and its matched pair alternated between blocks, so each singular filler (e.g, holiday vs foliday) would only occur 5 times during the experiment but one of the pair would occur in all 10 blocks.
All targets would appear with a different place holder sentence within each block: a place holder sentence and target pairing never occurred more than once. Trials within a block and blocks themselves were presented in a random order.

The focused attention manipulation came in the form of the instructions that participants saw, there were two sets: (i) Focused and (ii) Unfocused. Participants only ever saw one set. The instructions accounted for the pronunciation variation as 'mistakes'. The 'Focused' condition instructions alerted participants that possible mistake would always be in the final word and that changes could be small and would be sound based and took place at the start of the final word, although it did not divulge what sound these mistakes would affect. Whereas, in the 'Unfocused' condition the instructions simply stated that there could be some mistakes and did not emphasise which or where in the target word mistakes could occur. The instructions here diverged from Pitt and Szostak (2012) because of the differing task demands of having a place holder sentence which meant that we had to identify the final word as being the token that participants had to make a lexical decision on. Both sets of instructions can be seen in full in Appendix A.

Comprehension questions were included after 20% of trials, so that engagement with the task could be gauged throughout the task. These questions only followed trials which contained a 'word' filler target. The comprehension questions asked participants to select one of two choices: the target they just heard or a competitor word. The competitors were phonetically or semantically similar to the target word heard, for example, for one of the trials the filler target was "Drugs" and the competitor target for this trial was "Drums". There were an equal number of comprehension questions in each block.

The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. All materials were produced by a native British English speaker. The speaker was instructed to produce the materials in a naturalistic manner. Sentence place holders and target items were recorded separately and repeated until a delivery approximating natural speech was achieved. The sentential contexts were always produced with the token "pen" as the final referent, keeping the effects of co-articulation and prosody between the sentence and experimental target words constant throughout the experiment. Six of the fillers began with the same /p/ phoneme but filler trials were not included in the analysis. The sentential context was kept as recorded with only the final token excised. The target items were recorded using neutral sentence place holders to minimise list effects on the pronunciation of the target and keep the production replicating natural spoken language as much as possible. This context was then excised to leave just the target. A list version of the target was also produced immediately after the previous sentence version finished, this was so that we had more than one token to choose from for the experiment. The most natural sounding of the two target productions was chosen by the experimenter after testing with a place holder sentence. All recordings were saved in a mono 48kHz .wav format.

4.6.3 Apparatus and Procedure

The visual and audio stimuli were presented using DmDX software (version 5; Forster & Forster, 2013) on a PC and a 15 inch monitor set at a 1024x768 resolution.

Groups of up to 4 listeners were tested simultaneously across 2 rooms. The 2 computers in each room were separated by a divider meaning that participants could not see the computer screen of the other participant at any time during the experiment. If there were 2 participants in one session they were seated in different rooms. After reading an information sheet and filling in a consent form, listeners were seated at a computer and told to put on headphones that were attached to their computer. Although there could have been two listeners in a room simultaneously, there was unlikely to have been any noise distractions from the other participant due to the over-ear design of the headphone, which minimised ambient noise.

Participants then read through the practice trial instructions presented onscreen. These instructions matched the instructions of the main experiment and the distinction between focused and unfocused conditions was present in the practice trial instructions. The instructions asked participants to judge the final word of the sentence as either a word or not by pressing a key. Quick and accurate responding was also stressed. Following this they performed 4 practice trials, which did not vary across participants, these practice trials comprised of 4 filler trials taken from the main experiment and were designed as a familiarisation phase.

Trials started with a count down marker of "###", "##", "#"after which the trial would begin. This countdown marker was used so that participant's attention would be cued to focus on the auditory stimuli from the beginning. Locked the start of the place holder sentence, "+++++" was displayed on the screen as a visual cue for the duration of the place holder sentence and target word. After which participants had to select whether they thought the target was a word or not by pressing either the left or right 'CTRL' key. The 'Word' and 'Non-Word' responses were written on the side of screen relating to the key that needed to be pressed to select that answer. The position of the 'Word' response matched the dominant hand of the listener, as self-identified at the beginning of the study on the consent form. For example, if the participant was right handed then it would appear on the right hand side of the

screen. This meant that the 'Non-Word' response would appear on the side of the listener's weaker hand. Once a selection had been made or the trial timed out (3000ms), the next trial began automatically. After completing the practice phase, participants got the chance to ask any further questions of the experimenter. A large proportion of participants had noted the ambiguity and questioned how they should respond in the face of this. At this point the experimenter made clear that participants should go with their initial response as to whether the stimulus was a word or not and that there were no wrong answers.

Participants then viewed the repeated instructions as seen during the practice trials. There was an additional screen that alerted participant to the fact that comprehension questions followed some trials. Participants were told that they would have to choose from one of two answers that would be presented on screen and that they should press the 'CTRL' key that related to the side of the screen the answer they wished to select was on. If the answer was on the right side they should press the right 'CTRL' key. They then moved to the main experiment. The trial's here followed an identical structure to the practice trials described above. However, a comprehension question followed 20% of trials. After a selection was made in either a normal or comprehension question trial the next trial began automatically. There were breaks between blocks with a participant having to press a key to resume the experiment and move to the next block. The study took between 45-60 minutes to complete depending on the length of breaks taken by a participant.

4.6.4 Measures

The measures that were used were the proportion of word responses for each continuum point and the percentage of comprehension question that were answered correctly.

4.6.5 Analyses

We analysed participants' lexicality judgements. Our primary focus was the proportion of lexical responses and how this spread across the continuum when

broken down by Focus condition. All analyses relate only to the experimental target data; filler target trials were removed. Additionally, we excluded trials where participants did not make any selection and the trial timed out. This accounted for 0.9% of all trials. For the purposes of analysis, we created a new continuum variable: the continuum factor discussed below is not the absolute 5-step continuum from /s/ to /ʃ/ that is outlined above but a new 5-step continuum from non-word at point 1 to word at point 5. This was created by reversing the original /s/ to /ʃ/ continuum for the 'Chandelier' data, so that the /ʃ/ and, hence, the word end of the continuum were realigned with point 1, meaning that a continuum with 'word' end of the continuum for both experimental targets was created. For each trial, if the participant selected a word we coded this as 1 and if a non-word then this was coded as a 0. Due to our dependent variable being binomial (whether a participant judged a target as a word or not), we decided not to employ ANOVA analyses instead opting for a linear mixed-effects regression model with empirical logit transformed proportion data. This model was 'maximally specified' with both random intercepts and slopes, as well as their correlations varying by participants, as suggested by Barr, Levy, Scheepers and Tily (2013). The reasoning for the choice of an empirical logit transformation was that we expected that at the Continuum endpoints there would be a lot of either a lot of 0s but few 1s, or vice versa. When this occurs logistic regressions tend to have problems converging. This problem is minimised when an empirical logit transformation is employed. The predictors we used in the analyses were Focus (Focused and Unfocused) and Half (1 or 2) which were between participants and Target (Sandcastle and Chandelier) which was within participants.

The comprehension question data was used as a check throughout the experiment: If a participant was consistently answering comprehension questions wrong then this would question the validity of their data. For each comprehension question we coded 1 for a correct answer and 0 for an incorrect answer and then we created a percentage of correct responses for each participant.

4.7 Results

We first present the comprehension question results, as this affected the data taken forward into analysis of the main lexicality judgements. The results of the lexicality judgement analyses are presented following this. All lexicality judgments were analysed in R (R Development Core Team, 2014) using the lme4 package (Version 0.999999-0, Bates, Maechler & Bolker, 2014), p values were calculated using the lmerTest package (Version 1.2-0, Kuznetsova, Brockhoff & Bojesen, 2013).

4.7.1 Comprehension Questions

As described above, we wanted to check participants' answers to the comprehension questions to decide whether the rest of their data should be included in the analyses. The lowest comprehension question score was 94% of comprehension questions answered correctly. On this basis all participants' data was included in the main analyses.

4.7.2 Proportion of Lexical Responses

Central to our predictions was the effect of focused attention on the proportion of word responses made by participants. Figure 4.2 shows there were no reliable differences (average= 0.05) between the proportions of word responses following either focused or unfocused instructions. There was slightly more variation at the non-word end of the continuum until the midpoint at Point 3, with the largest difference of 0.12 at continuum point 2. At Points 4 & 5 near identical proportions (a difference of 0.02 or below) of word responses were seen. We used empirical logit transformed proportion data and a linear mixed model, as described above, with Focus and Target as sum coded predictors and this confirmed there was no effect of Focus, ($t < 1$).

However, as seen in Figure 4.3, there was notable variation (average= 0.3) between the two target words: listeners were much more accepting of pronunciation variation in the 'Chandelier' target, with it having a higher proportion of word responses at every continuum point bar point 5. This pattern was especially prevalent in the middle continuum points (2 and 3) with both these continuum points having a difference between target words of over 0.5.

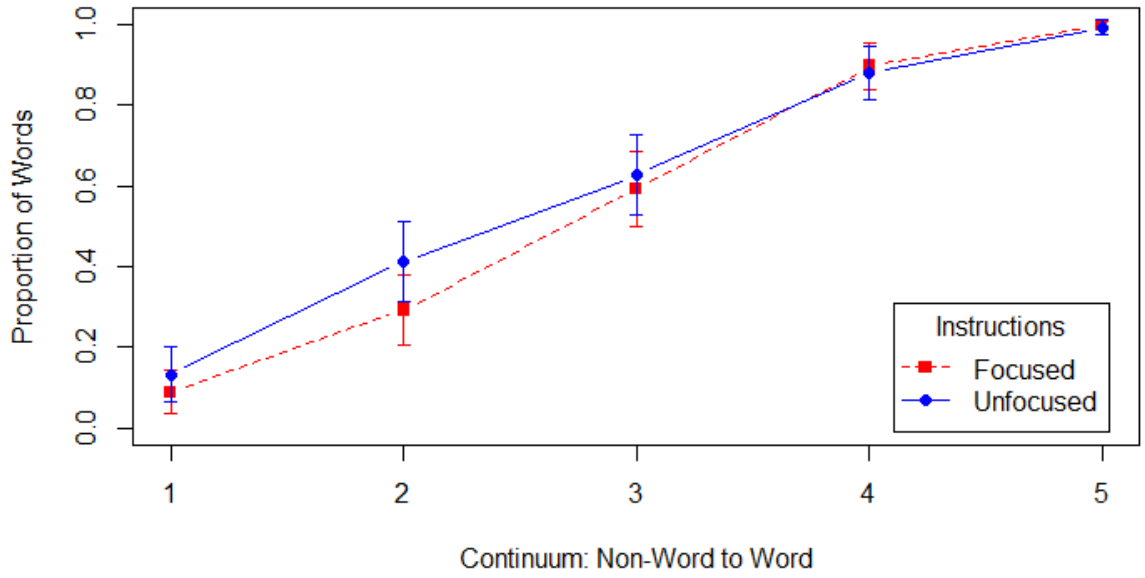


Figure 4.2- The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused).

At the endpoints of the continuum (1 and 5) there was convergence between targets. This was expected as the target word heard either had no pronunciation variation and represented a normal production of the word (5) or the greatest pronunciation variation with the opposite end of the fricative continuum present (1), meaning that these should have been the most clear distinctions between a word and non-word target for a listener. Unsurprisingly, a reliable target effect was observed (β= -0.92, SE=0.15, t=-6.25, p <0.001).

Figure 4.4 shows the interaction between Target and Focus, breaking each target word down by instruction type (Focused & Unfocused). There was increased differences seen between the focus conditions for 'Chandelier' (average= 0.09): The Focused instruction condition reduced the proportion of word responses at the 'non-word' end of the continuum. The largest magnitude was at points 2 and 3 and to a lesser extent point 1. At the 'word' end, from point 3 onwards the responses in are matched at above 95% word responses. The 'Sandcastle' target shows very little variation between focused conditions at any point. There is no interaction effect between Focus and Target (t<1).

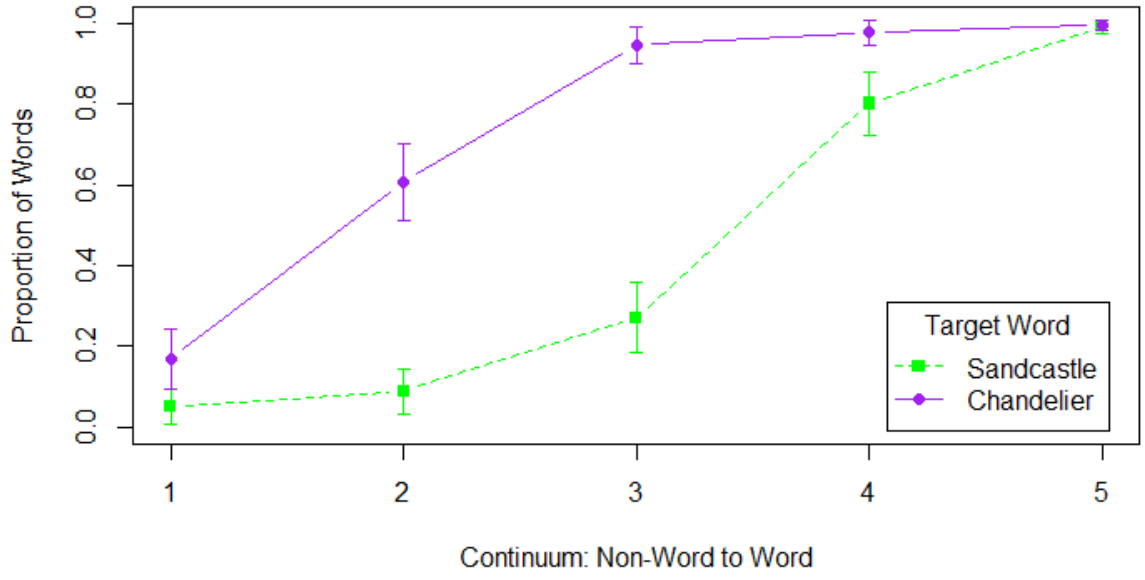


Figure 4.3- The proportion of word responses for each Continuum Point (Non-Word to Word) by Target word (Sandcastle and Chandelier).

The experiment was long (50-60 minutes) and repetitive (300 trials), we wanted to check whether participants were paying less attention towards the end of experiment and whether this could possibly mask an effect. To guard against this, we created a new predictor, 'Half' which coded whether a trial occurred in the first or second half of the experiment. Running a separate linear mixed effects regression

model still with empirical logit transformed proportions with half as a predictor revealed no effect (t <1.5).

4.8 Discussion

The aim of the current study was to show a replication of the focused attention effect observed in Pitt and Szostak (2012) but there was no reliable difference seen by focus condition. However, the small differences seen between conditions were in the direction that we would have predicted, with focused attention bringing the proportion of word responses down which was encouraging for future studies.

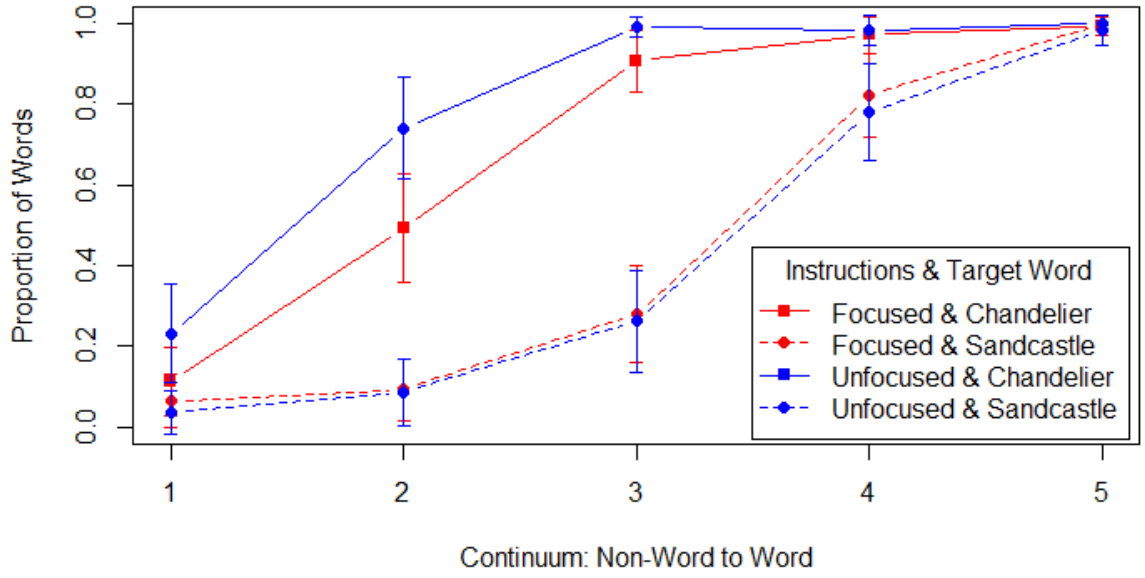


Figure 4.4- The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused) and then by Target word (Sandcastle and Chandelier).

We cannot directly compare to the Pitt and Szostak study as the current study employed an updated paradigm. It is unlikely that putting the target words following short, context neutral sentence stem should be responsible for the lack of variance between the focused and unfocused conditions. In the current study, the 'chandelier' target produced comparable results to those seen for the same target in

Pitt and Szostak, this suggests that even with the introduction of a neutral sentence placeholder that results need not be too heavily impacted or divergent from in the single word lexical decision task. However it is possible that due to the participants having to comprehend preceding material there may have been a reduction in their attention to the sentence final word when compared to the Pitt and Szostak study. In the Pitt and Szostak paradigm, participants knew that they were only being presented with single words, meaning that the next word was always judged as a crucial target, whereas, in the current paradigm it may not have been apparent which was the critical target word for the participants to respond to whether it was lexical or not.

A second and complimentary reason that we propose as a possible driver for the lack of replication for an attentional effect stems from using a compound noun, 'sandcastle', as it is likely that this may have caused some confusion in participants when responding under time pressure. A third complimentary reason is the inclusion of a second /s/ phoneme in 'sandcastle'. In the instructions we signposted participants to the word and the location of the variation in the target word, so although they may have been aware of the word initial phoneme variation the second competing phoneme sound could have caused confusion. Whether any confusion would have been likely to cause participants to rate the target as less 'word' like is hard to know but it is a weakness in the paradigm that needs to be addressed in the following studies. Interestingly, participants seemed to judge pronunciation variation as less 'word' like when it was housed in the /s/ predicted target, 'sandcastle'. In both the pre-test and main experiment participants rated the /ʃ/ expected target, 'chandelier' as notably higher for lexicality judgements. This suggests that participants may be more comfortable with variation in the place of /ʃ/ over /s/. Although the differences seen here could be attributed to other factors linked to the target word selection as outlined above, it will be interesting to see if this pattern is repeated in the following studies.

Taken together these findings do not support an attentional effect as predicted but as we have detailed there are a number of reasons as to why the results seen above may not prove conclusive. These stem from the /s/ phoneme target word selection and changes to the paradigm that may have produced a reduced or even lack of effect due to the impacting upon participants decision making process with that seen in Pitt and Szostak (2012). These findings from the current study provide a useful starting point to build on with the next study in maximising the efficacy of the paradigm.

4.9 Experiment 3

Following Experiment 2 we continued to investigate how focused attention impacted upon participants' lexical decisions to target stimuli in a sentence based speech perception paradigm. This method for focusing attention was adapted from Pitt and Szostak (2012). Their study produced an attentional effect: participants who saw the focused instructions produced a lower proportion of word responses during the lexical decision task. We had predicted the same attentional effect to be observed for the previous study but there was a lack of variance between attentional conditions.

Our interest in attention stems from the central focus of the current series of experiments, which is testing the attentional account of disfluency as outlined in the chapter overview. However we first wanted to establish that we could replicate an attentional effect using our materials and sentence based paradigm but this was not the case. We suggested two explanations that may have been responsible for the lack of attentional effect: Firstly, the introduction of a placeholder sentence creating different task demands and expectation than those participants encountered in Pitt and Szostak (2012). Secondly, our selection of 'sandcastle' as a target word, as this target showed a divergent pattern of results to the 'chandelier' target, which had previously been used in the Pitt and Szostak study (2012). We noted that both it

being a compound noun and the inclusion of a second /s/ sound may have caused additional confusion during the lexical decision task.

In the current experiment, we repeat the previous paradigm whilst making necessary changes to counter the repetition of a second /s/ phoneme that may have been a weakness in the choice of previous target words. It was only following the current study that we identified that 'Sandcastle' being a compound noun may have negatively influenced participants' lexical decisions. Therefore this issue is not addressed in the current study. Instead, we employed another compound noun, 'Sandpit' for the current paradigm. Our primary aim for the current study was again to replicate the attentional effect observed in Pitt and Szostak (2012) with our sentence level paradigm and updated materials before the addition of disfluency. Our predictions remain the same as in the previous experiment: Based on the attentional effect seen in Pitt and Szostak (2012) we would expect an increased reliance on bottom-up phonetic information for those participants who saw the 'focused' instructions. We would expect this to result in an increase in the proportion of 'non-word' responses in comparison to the 'unfocused' instructions. Hence, attention would modulate the participants' rating of lexical acceptability for a word. An additional lexical bias would be expected following the 'unfocused' instructions, as there would be nothing driving increased attention to the fine-grained acoustic information contained in the incoming speech stream.

4.9.1 Target Word Selection

As noted above in Experiment 2's results section, the /ʃ/ initial target word 'Chandelier' showed a similar pattern of word responses across the continuum to those seen in the results of Pitt & Szostak (2012). Although there was a raised proportion of word responses from continuum point 3 onwards. In contrast, 'Sandcastle' showed a pattern vastly different from both the pattern seen in Pitt & Szostak (2012) and from the 'Chandelier' target word. There were equivalent proportions seen for each continuum point for both Focused and Unfocused

conditions. We suggested that a reason for this differing pattern and lack of focus effect could be the repetition of a second /s/ phoneme following the word initial occurrence is likely to have complicated the lexicality judgement for listeners.

We addressed this potential flaw in the current follow up study by replacing our /s/ variant target, 'Sandcastle' for the related 'Sandpit'. Crucially, there was no reoccurrence of any related /s/ or /ʃ/ following the word initial variant. Additionally this target word met other necessary criteria, as outlined in that target word selection section in Experiment 2: It formed a word at one end of the continuum ('Sandpit') and a non-word at the other end ('Shandpit'); it is close in phonemes (7) to Chandelier (8); it contained the same vowel sound /æ/ as heard in 'Chandelier'. The same continuum used in the previous experiment was employed for the current study due to the phoneme variation still occurring word initially. As the current experiment used the same continuum and the new /s/ variant, 'Sandpit', was a modified version of the previous target, it was decided that further norming studies were not a necessity.

The new target was produced in a new session but it was recorded at the same studio with the same engineer and employed the same native English speaker who produced the materials in Experiment 2. Again, the speaker was instructed to produce the materials in a naturalistic manner and multiple variants were recorded until a delivery approximating natural speech was achieved. The process followed was identical to that used to record the target words in the previous experiment: Namely using neutral sentence place holders to minimise list effects on the pronunciation of the target and to keep the production replicating natural spoken language as much as possible. This context was then excised to leave just the target. A list version of the target was also produced immediately after the previous sentence version finished. This allowed us to have multiple productions to choose from when selecting for the current experiment. The most natural sounding of the two target productions was chosen by the experimenter after testing with a place holder sentence. All recordings were saved in a mono 48kHz .wav format.

4.9.2 Participants

A total of 25 students from the University of Edinburgh participated for a reward of £6.50 upon successful completion of the current and another short study. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties. Participants who had taken any of the pre-tests for the previous experiment or experiment 2 itself were excluded from taking part in the current study.

4.9.3 Design and Materials

The continuum, filler targets and sentence place holders were reused from the previous experiment; details on these are given in the previous chapter. They are recapped in short below. Each trial was made up of a place holder sentence (e.g, '*I had lost the…*') followed by a target word that participants then had to make a lexicality judgement on. There were two types of target word: experimental targets and filler targets. The experimental targets varied from the previous experiment, with 'Sandpit' replacing 'Sandcastle'. 'Chandelier' was repeated. The experimental targets each had 5 variants that corresponded to a change in their initial phoneme sound for each place of continuum described above. At one end of the continuum (as in normal productions) each experimental target was a word, whilst the same target was made into a non-word at the other end of the continuum: 'Sandpit' became 'Shandpit'; 'Chandelier' became 'Sandelier'.

The fillers and design employed matched those detailed in the design and materials of Experiment 2. Comprehension questions were again presented after 20% of trials.

4.9.4 Apparatus and Procedure

The apparatus and procedure followed those detailed above for Experiment 2.

4.9.5 Measures

The measures that were used were the proportion of word responses for each continuum point and the percentage of comprehension question that were answered correctly.

4.9.6 Analyses

We analysed participants' lexicality judgements in the same way as in Experiment 2. Our primary focus was the proportion of lexical responses and how this looked across the continuum. All analyses are only on the experimental target data, filler targets are removed. We excluded trials where participants did not make any selection and the trial timed out, this accounted for 1.4% of all trials.

The continuum factor discussed below is not the absolute 5-step continuum from /s/ to /ʃ/ that is outlined above but a new 5-step continuum from non-word at point 1 to word at point 5 as in Experiment 2. The dependent variable was again binomial (whether a participant judged a target as a word or not), therefore, as in Experiment 2 we opted for a linear mixed-effects regression model with empirical logit transformed proportion data. This model was 'maximally specified' with both random intercepts and slopes, as well as their correlations varying by participants, as suggested by Barr, Levy, Scheepers, and Tily (2013). The reasoning for the choice of an empirical logit transformation is detailed in Experiment 2. The predictors we use in the analyses are Focus (Focused and Unfocused) which is between participants and Target (Sandcastle and Chandelier) which is within participants.

The comprehension question data was used as a check throughout the experiment: If a participant was consistently answering comprehension questions wrong then you would question the validity of their data. For each comprehension question we coded 1 for a correct answer and 0 for an incorrect answer and then we created a percentage for each participant, based on the number of correct responses.

4.10 Results

We first present the comprehension question results, as affected the data taken forward into analysis of the main lexicality judgements. The results of this lexicality judgement analyses are presented following this. All lexicality judgments were analysed in R (R Development Core Team, 2014) using the lme4 package (Version 0.999999-0, Bates, Maechler & Bolker, 2014), p values were calculated using the lmerTest package (Version 1.2-0, Kuznetsova, Brockhoff & Bojesen, 2013).

4.10.1 Comprehension Questions

A low percentage of comprehension questions answered correctly would indicate that a participant may have struggled with the task or that they were not giving the task their full attention. The lowest comprehension question percentage for the current study was 89%. Again, on this basis no data was excluded from the analyses for this reason.

4.10.2 Proportion of Lexical Responses

Our primary focus is on the effect of focused attention across the continuum and how this influenced the proportion of 'word' responses. We predicted that in the 'Focused' attention instruction condition we would see a reduction in the proportion of word responses. Figure 4.5 showed that this pattern of results is not seen here, as there was no reliable difference between attention conditions (mean= 0.04). The direction of the pattern was the reverse to what we predicted seen with the 'Focused' condition having a matched or higher proportion of 'word' responses at every continuum point. Using empirical logit transformed proportion data and a linear mixed model, as described above, with Focus and Target as sum coded predictors confirms there is no effect of Focus, ($t < 1$).

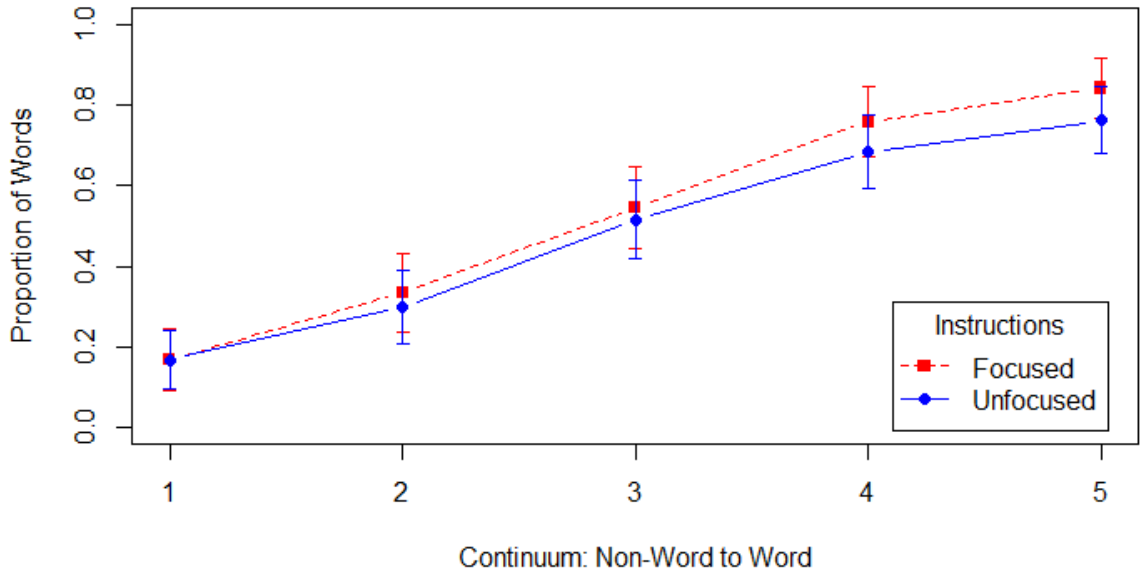


Figure 4.5- The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused).

In comparison to Experiment 2, listeners were less tolerant of fricative variation, especially towards the 'word' end of the Continuum: Both attention conditions had under a 0.85 for proportions of 'word' responses at Continuum point 5, whereas, in Experiment 2 Continuum point 4 & 5 were all rated as above 0.8 for proportions of word responses.

We changed the /s/ initial target word to 'Sandpit' for the current study. The pattern of the proportion of word responses by Target can be seen in Figure 4.6. This graph showed that large differences between target words still remained in the current study (mean= 0.43): 'Chandelier' again follows the pattern seen in Experiment 2 and Pitt & Szostak with 'Sandpit' having a lower proportion of 'word' responses across the continuum. The gap is particularly pronounced in the middle continuum points with an average difference of 0.6 between target words. The linear mixed model confirmed a significant target effect (β= -0.93, SE=0.11, t=8.1, p <0.001).

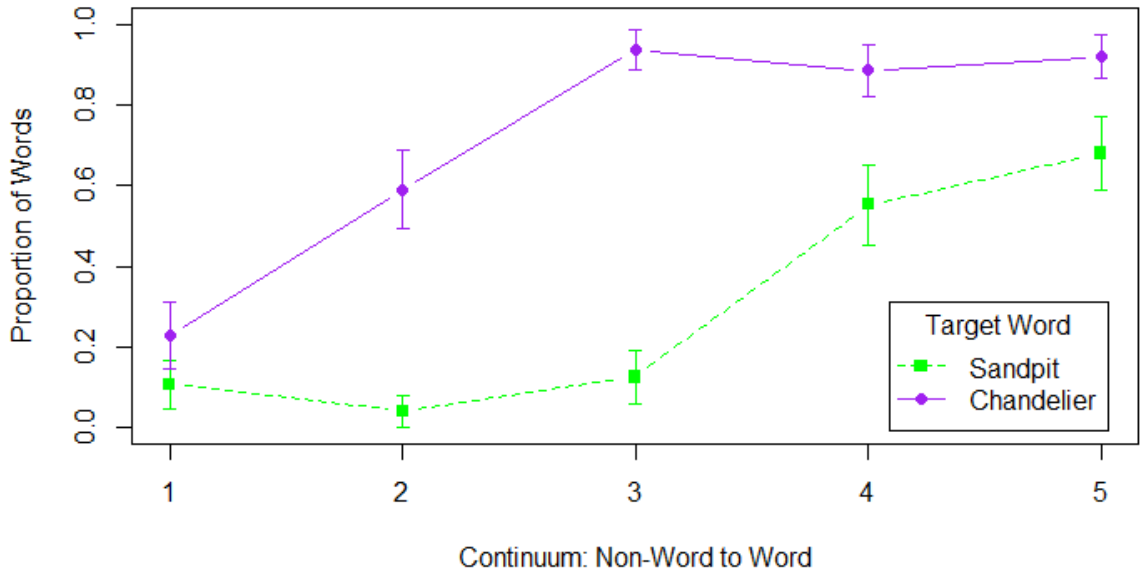


Figure 4.6- The proportion of word responses for each Continuum Point (Non-Word to Word) by Target word (Sandpit and Chandelier).

In comparison to Experiment 2, the participant's tolerance of fricative variation across the continuum for 'Chandelier' was equivalent with the same pattern seen. However, for 'Sandpit' there were notable decreases in tolerance of fricative variation at Continuum points 4 and 5 in comparison to Experiment 2 and 'Sandcastle': At 5, a large negative magnitude, bringing it down from a 0.99 proportion of word responses to a 0.68.

In experiment 2, when the Focus condition was broken down by Target, there was more variation seen between the 'Focused' and 'Unfocused' conditions for 'Chandelier' and no variation seen for 'Sandcastle'. There is a different pattern for the current study, as seen in Figure 4.7, with some differences observed between the Focus conditions for each target. However, this was still negligible with an average difference in proportion of word responses of 0.04 between Focus conditions for 'Chanderlier' and 0.08 for 'Sandpit'. There was no focus by target interaction, ($t < 1$).

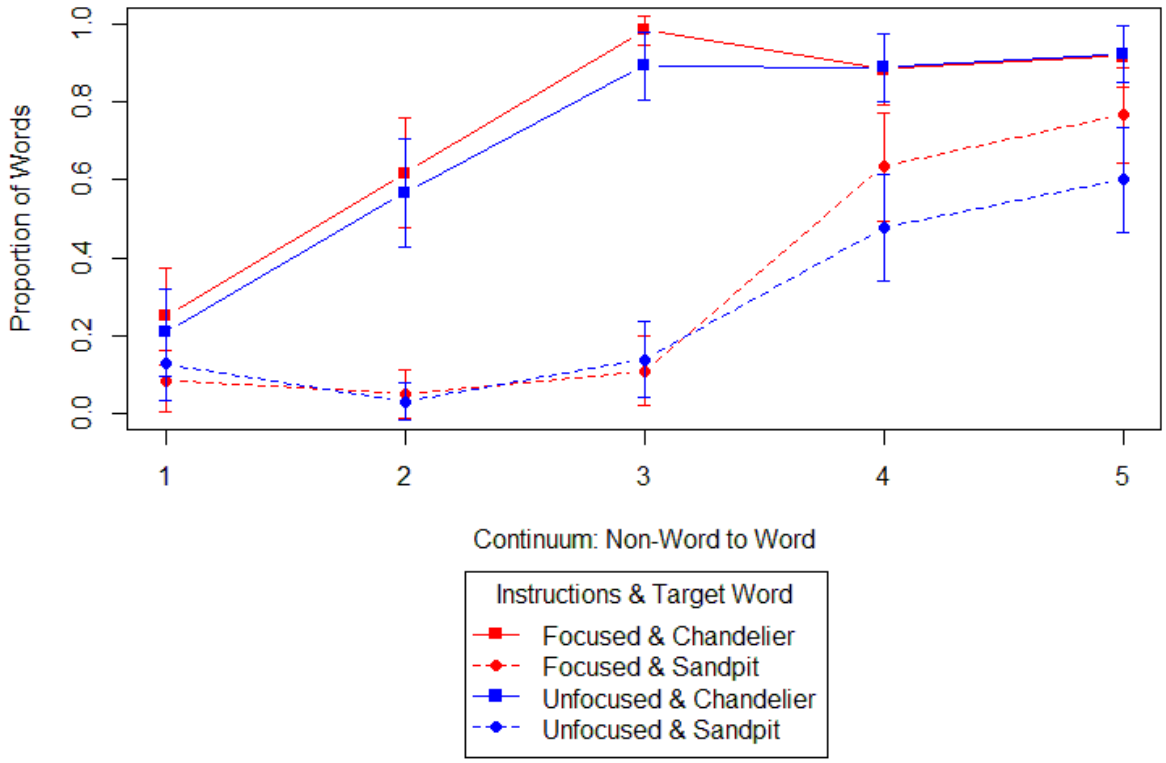


Figure 4.7- The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused) and then by Target word (Sandpit and Chandelier).

The experiment was long, with the duration (50-60 minutes) matching the previous study. It was also repetitive with 300 trials. Again, we were concerned that a loss of attention from participants would have resulted in suboptimal performance that could have impacted upon the results. To guard against this, we again included the predictor, 'Half' which coded whether a trial occurred in the first or second half of the experiment. Running a separate linear mixed effects regression model still with empirical logit transformed proportions with Half as a predictor revealed no effect ($t < 1.1$).

4.11 Discussion

The aim of the current study was to show a replication of the focused attention effect observed in Pitt and Szostak (2012) but again there was no reliable difference

seen by focus condition. It is worth repeating that the current study, as with experiment 2, is not directly comparable with Pitt and Szostak (2012) due to the addition of a neutral sentence place holder here, compared to a single target item in the original study. As suggested in the previous study, it is possible that due to the participants having to comprehend preceding material there may have been a reduction in their attention to the sentence final word when compared to the Pitt and Szostak study. In the Pitt and Szostak paradigm, participants knew that they were only being presented with single words, meaning that the next word was always judged as a crucial target, whereas, in the current paradigm it may not have been apparent which was the critical target words to respond to on whether it was lexical or not. However the current pattern of responses across the continuum for the 'Chandelier' target followed a similar pattern to that observed in Pitt and Szostak (2012) which does not support the use of a sentence placeholder creating vast variance from the single word paradigm.

The results did show a roughly similar pattern of results to those seen in Pitt and Szostak paper with differences in the proportion of lexical judgements between continuum points and a strong tendency for increasing proportion of word response at the 'word' end of the continuum. In the current study we removed the second /s/ phoneme that featured in 'Sandcastle' in experiment 2, instead replacing it with 'Sandpit'.

We suggested in the previous study that the use of a compound noun may have created confusion for participants, as 'Sandpit' could be broken down into constituent parts ('Sand' + 'Pit') that are equally lexical. However we did not control for this in the current study as this possible weakness only became apparent following the results. The results seen here provided some support for this suggestion as even at the 'word' end of the continuum (Point 5) there was a relatively low proportion of lexical judgments (0.68) for the stimuli that had no phoneme variation. In the further studies in the current series, we will avoid

compound nouns for this reason. As we have only had two target words the results are susceptible to any variance that may be associated with individual target words.

Our lack of replication effect may have been from other factors that we cannot investigate from the current results such as the acoustic nature of stimuli from Pitt and Szostak's or differences between British and American English or sensitivity to variation that may vary between groups of participants from these respective countries. If our continuum was more natural sounding than Pitt and Szostak's then this could have influenced participants' sensitivity to the phoneme variation heard.

There was some similarities that held between the current study and experiment 2. Again 'Chandelier' received a significantly higher proportion of word responses in comparison to 'Sandpit', especially prevalent in the middle points of the continuum. This supports a similar pattern of results seen in the previous experiment, where participants also judged pronunciation variation as less 'word' like when it was housed in the /s/ predicted target, 'sandcastle'. It lends weight to the proposal that participants may have been more comfortable with variation in the /ʃ/ target word over the /s/ target. However, this difference between target words could have been driven by the /s/ target being a compound noun or another form of variance that is associated with this specific target word, as detailed above. We cannot know if this was the case but by controlling this factor going forward we can rule it out. Additionally we cannot compare this to the Pitt and Szostak results as they are not broken down by target, so there may have been differences between their targets.

The findings of Experiment 2 and the current study have not supported the attentional effect we predicted but we have gained a better understanding of the paradigm and highlighted areas that we can strengthen in the remaining studies of the current series. The results have provided additional insight into how we should conduct the next study to create a paradigm that minimises the possible weakness in the /s/ target word identified here.

CHAPTER 5

Experiment 4

5.1 Introduction

The lack of predicted results in Experiment 3 meant that we had to further adapt our experimental paradigm to maximise the possibility of generating an effect of focused attention, as seen in Pitt and Szostak (2012). It was crucial to first establish that we could get an effect of focused attention before the addition of disfluency into the paradigm; otherwise effect attribution may have become complicated due to crossed conditions. There were a number of paradigmatic lessons that we learnt from Experiment 2 and 3. In both previous experiments our/ʃ/ initial target word 'Chandelier' showed a similar pattern of word responses across the continuum as seen in the results of both experiments in Pitt and Szostak (2012), although with a raised proportion of word responses from continuum point 3 onwards.
In comparison, the /s/ initial target word in the previous experiments, 'Sandcastle' and 'Sandpit' respectively, revealed large differences to both the /ʃ/ initial target word featured in both studies ('Chandelier') and the general pattern for word initial fricative variation observed in both experiments in Pitt and Szostak (2012). However, it is important to note that in Pitt and Szostak (2012) there were no analyses of the data broken down into the constituent target words: The results were collapsed across /s/ and /ʃ/ target words. So there may have been similar variation between /s/ and /ʃ/ alternatives but we have no way of knowing or comparing to our own results. Similarly, for the previous experiments (2 & 3) the addition of the /s/ target word data to the/ʃ/ variant decreased the average proportion when breaking the data down by Focus condition (the data graphed in Pitt & Szostak, 2012).

As noted above, in Experiment 2 'Sandcastle' showed vastly lower proportions of word responses at Continuum medial positions in comparison to /ʃ/ target. At the endpoints, there was convergence with the 'Chandelier' target word proportions. We

suggested that a reason for this differing pattern of responses by target word and lack of focus effect could be that the repetition of a second /s/ phoneme following the word initial occurrence is likely to have complicated the lexicality judgement for listeners. In Experiment 3, 'Sandpit' again showed much lower proportions for word responses at each Continuum point compared to 'Chandelier'. The largest differences were again seen continuum medially but with reduced convergence at the continuum end points. The 'Sandpit' proportion values were lower than for 'Sandcastle'. There was no repetition of either /s/ or /ʃ/ in 'Sandpit' but we suggested that the compound noun classification of this target could have caused ambiguity for the lexical judgements of participants, especially under time pressure; they may have been unsure whether they had to judge 'Sandpit' as a whole or just 'pit', with the latter being a perfectly good word by itself.

Due to the lower proportions across /s/ target words observed in both experiments, we suggested there could have been differing tolerances of fricative variation between the /s/- and /ʃ/-containing target words, although the limited number of target words meant we cannot generalise with certainty to this extent. The reduced proportions for /s/ targets may have been linked to anomalies with the target words tested in the previous experiment, especially as they were so closely linked, with the fricative variation occurring with same phonemic context, 'sand', in both. It is also possible that differences in the nature of the fricative variation between studies and the sensitivity to this variation exist between American English and British English listeners: either reason could cause different expectations between the sets of listeners and could explain the different results seen here. The different task demands and, hence, changes in the instructions mean that there were unavoidable differences from the Pitt and Szostak (2012) study. Their study focused on lexical judgements for single words only and our focus in the previous and current studies extended their arguments to a sentence level by using place holder sentences.

The current study set out to use the knowledge gained during the previous experiments to create a revised paradigm that used new materials addressing the

issues encountered in the previous experiments. This meant making certain changes: Firstly, moving the fricative variation from a word initial to a word-medial position. Our motivation for this change was that Pitt and Szostak (2012) saw the largest effect of focused attention when the fricative variation occurred in a word-medial position. We had previously opted to use a word initial position for the Continuum as it was easier to create matched stimuli for this position compared to a word medial position. A word initial location did not require the fricative variation being implanted into the middle of a word, hence, not requiring a word being cut into parts. Additionally in previous studies a word initial location has led to robust results for speech perception studies (e.g., Ganong, 1980; Pitt & Szostak, 2012), including in a sentence based paradigm (Connine, 1987).

We selected our target words to accommodate the new word medial continuum position. We also controlled the target words for instances of /s/ or /ʃ/ phonemes that occurred in the remainder of the words. We did not want any further repetitions of these phonemes to complicate lexical decisions for participants. The target words were also not compound nouns, as again this could cause confusion about the entity being judged. The criteria for the selection of target words are discussed in more detail below. We changed our sentence place holders to make them compatible with the new target words. We also increased the percentage of filler trials from 60% to 86%, to match the percentage of experimental items to filler items seen in Experiment 2 in Pitt and Szostak (2012). Additionally, we reduced the number of filler targets, so that each participant would hear the same number of repetitions of each filler target and experimental target. Previously participants had heard each experimental target more than any of the filler targets which may have allowed them to be able to identify experimental targets or attach extra importance or have developed a strategy to deal with these targets. Therefore, by matching the number of repetitions for filler targets and experimental targets we can better disguise the experimental targets. The constant number of repetitions of each target type meant that participants were less likely to be able to identify experimental targets as the focus of the study.

The primary focus of the current study was to replicate the lexically biased attention effect seen in Pitt & Szostak (2012). We continued to manipulate focused attention in the form of the instructions seen by participants. The 'Focused' condition drew participants' attention to the word and the location within a word where the phonemic variation occurs. This was not the case in the 'Unfocused' condition where listeners were only instructed that variation could be present. This did not guide their attention to the final word or the word initial phoneme. Our predictions still centred on those participants that received the 'focused' instructions being less accepting in their judgements of the manipulated pronunciations as words. If this effect were to be found, we would be interested to see how it varied across the continuum, how it compared to the word-initial continuum, the influence of the variance on individual target words and the size of the effect relative to Pitt and Szostak's (2012) demonstration of an attentional effect. These questions could help us strengthen our knowledge of the paradigm and understand how to improve it. Our ultimate aim was to be able to extend the paradigm to include disfluency and compare the results to the 'focused' attention effect we hoped to observe here allowing us to investigate the attentional account of disfluency processing.

5.2 Target Word Selection

Due to the new word-medial position of the continuum in the target words, we had to select new target words to accommodate the fricative variation location. For the current study we took the late medial target-word pair used in Pitt and Szostak (2012), 'Impressive' and 'Condition'. Pitt and Szostak (2012) had two medial positions: early and late. We opted for late medial, as this word-position produced the largest effect of focused attention. However each target word had different phonemes surrounding the fricative (Impressive: Pre-/ɛ/and Post-/ɪ/; Condition: Pre /ɪ/ and /ə/). This word pair as lexical entries had effectively been pre-tested in the Pitt and Szostak (2012) studies and focused attention had been shown to exert an effect of reducing lexical judgements. The target word pair were similar for syllable

structure (both are trisyllabic) and matched in number of phonemes (8). They also met the basic criterion that at one end of the fricative continuum a 'word' was created and at the other end a 'non-word': For the /s/ target variant, 'Impressive' to 'Impreshive' and for the /ʃ/ variant, 'Condition' to 'Condision'. The final duration of 'Impressive' was 849ms; 'Condition' was shorter at 656ms.

5.2.1 Continuum Creation

First, we obtained the /s/ and /ʃ/ tokens that form the endpoints of the continuum by recording productions in a /ə/ context housed within words. These phonemes were then created by excising this vowel sound. The /s/ and /ʃ/ phonemes were produced in a number of words, with the clearest example of each chosen to be used. The fricatives were then matched for duration and loudness.

A 15 point word medial /s/ to /ʃ/ continuum was created by digitally blending the two phonemes in varying proportions. This continuum creation method is adapted from Pitt and Szostak (2012). The duration for all word medial fricative stimuli was 144ms. This duration was close to the 134ms medial fricative variation seen in Pitt and Szostak (2012). We aimed for a 5 step continuum with points 1, 3 and 5 set as fixed proportions: Point 1-100% /s/; Point 3- 50% of each phoneme and Point 5- 100% /ʃ/. Point 1, 8 and 15 from the extended continuum corresponded to Points 1, 3 and 5 in the 5 step continuum. The 12 intermediate steps (6 each side of the midpoint) in the continuum were made by blending together the phonemes in different proportions in 5% intervals. This created 6 steps on the majority /s/ side from which to select Point 2 and 6 steps on the majority /ʃ/ side of the continuum from which to select Point 4.

The materials used to create the continuum were recorded at a University of Edinburgh studio facility by an engineer with the author present. All materials were produced by a native British English speaker. The speaker was instructed to produce the words containing the /s/ and /ʃ/ tokens in a natural manner. The recording of the

continuum took place during the recording session that produced the rest of materials detailed below. All recordings were saved in a mono 48kHz .wav format.

5.3 Norming Studies

The current study was designed to elicit theoretical differences in lexical responses along a word to non-word continuum using /s/ and /ʃ/ phonemes in a pair of target words: 'Impressive' & 'Condition'. Therefore, we needed to make sure that the steps along our continuum both: (i) differed enough from one another to speak to our predictions and (ii) did not occupy the extremes of proportions, as if the points were rated as either 1 or 0 or values which were close to them this could mean responses would be subject to ceiling or floor effects.

5.3.1 Pre-Test 1

This pre-test measured participants' lexical judgements to each continuum point on the 15 step version created. The pre-test data was used to decide which phoneme variation would be selected from the intermediate points of the 15 step continuum to feature in the final 5 step continuum. It also tested if participants were sensitive to differences between continuum points. It is worth noting that for the current pre-test the continuum being discussed is in absolute terms: from Point 1 which was 100% /s/ to Point 15 which was 100% /ʃ/. This is not to be confused with the Continuum predictor employed in previous Results sections, which is a relative continuum that runs from 'Non-Word' to 'Word'.

We predicted that participants' responses would decrease from the 'word' end of the continuum and should decrease further with increasing fricative variation towards the 'non-word' endpoint: This means that at each end of the continuum we would expect different patterns between the target words: At Point 1 which is 100% /s/ we would expect to see high proportions of word responses for 'Impressive' but not 'Condition' and the inverse would be true for Point 15 which is 100% /ʃ/. The pre-test

employed single words that were presented individually, rather than in sentential context. The current paradigm was a replication of Pitt and Szostak's (2012) paradigm but our primary focus was pre-testing the continuum for the main experiment. However the results here would create an interesting replication of Pitt and Szostak's study (2012).

5.3.2 Participants

A total of 10 students from the University of Edinburgh psychology community participated in the experiment for a reward of £4.50 upon successful completion of the experiment. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties.

5.3.3 Materials and Design

Each of the 174 trials was a single word target that participants had to judge either as a 'word' or 'non-word'. There were two types of target: Experimental Targets and Filler Targets. The 30 Experimental Targets were the 15 steps of the continuum described above spliced into the target words, 'Impressive' & 'Condition'. This created a continuum of target words: each target with each of the 15 steps from the word-medial phoneme continuum. The 12 Filler targets were made up of 6 matched word and non-word pairs. Each bisyllabic or trisyllabic word had a matched equivalent that manipulated a medial phoneme to make it into a non-word ('Holiday' and its matched non-word equivalent was 'Holinay'). All manipulated phonemes differed only in either place or manner of articulation. All fillers were repeated 12 times, creating 144 filler items. All items were randomly presented in a single block. The experimental targets made up 17% of trials with fillers presented for the remaining 83% of trials. This was similar to the percentages of experimental targets (14%) and fillers (86%) seen in Experiment 2 in Pitt and Szostak (2012). The instructions for the current study were equivalent to the 'unfocused' condition in the previous experiments because they did not draw attention to the location within the

word where the fricative variation occurred. They also differed from previous experiments as they acted on a word rather than a sentence level.

All materials were produced by the same native British English speaker as in experiments 2 and 3. The speaker was kept constant so there was consistency across experiments and that any different effects could not be attributed to a change in speaker. The speaker was instructed to produce the materials in a naturalistic manner. The target items were recorded using neutral sentence place holders, such as "*I want the*…". This method of recording minimised list effects on the pronunciation of the target and kept the production replicating natural spoken language as much as possible. This sentential context was then excised to leave just the target. The method and process of recording was the same as in Experiments 2 and 3. The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. All recordings were saved in a mono 48kHz .wav format.

5.3.4 Apparatus and Procedure

The visual and audio stimuli were presented using DmDX software (version 5; Forster & Forster, 2013) on a PC and a 15 inch monitor set at a 1024x768 resolution. Up to two listeners could be tested simultaneously. The 2 computers in the lab were separated by a divider meaning that participants could not see the computer screen of the other participant at any time during the experiment. After reading an information sheet and filling in a consent form, subjects were seated at a computer and told to put on headphones that were attached to their computer. Although there could have been two listeners in a room simultaneously, there was unlikely to have been any noise distraction from the other participant due to the over-ear design of the headphone, which minimised ambient noise. Participants then read through the instructions presented onscreen. The instructions asked participants to judge the speech as either a word or not by pressing a key. Participants then moved to the data collection phase of the experiment. The structure of the trials matched the previous

experiments (2 & 3) apart from upon the display of "+++++" it was the target word that began playing instead of a sentence place holder. Participants still had to select whether they thought the target was a word or not by pressing the relevant left or right 'CTRL' key. The time out value was reduced to 1800ms from the 3000ms used in Experiments 2 and 3 because of the shorter duration of the audio stimuli being presented. Only a single word was presented, so they did not have to process any other auditory information before answering, this contrasted with our previous studies in which participants had to listen to and process the sentence contexts before hearing the target and responding. The study took approximately 15 minutes to complete.

5.3.5 Measures

The measure that we used was the proportion of word responses for each continuum point.

5.3.6 Analyses

We analysed participants' lexicality judgements for each continuum point. All analyses were run only on the experimental target data; filler targets were removed. We excluded trials where participants did not make any selection and the trial timed out; this accounted for 4% of all trials. For each trial, if the participant selected a word we coded this as 1 and if a non-word then this was coded as a 0.

5.3.7 Results

The primary goal of the pre-test was to select the two intermediate continuum points (2 & 4) to complete our 5 step continuum. In the intermediate positions our criterion were stimuli that: elicited strong lexical biases and covered a range of fricative variation, so that our proportion of 'word' responses would cover a range of values from 0-1 and not be subject to floor or ceiling effects. Table 5.1 shows the mean proportion of 'word' responses broken down by the two target words being tested. From this graph it can be seen that the 15 step continuum is not balanced by target

word. Participants were more tolerant of the fricative variant in the /ʃ/ based target word, 'Condition' than for the /s/ based, 'Impressive'. Table 5.2 shows the mean proportion of 'word' responses calculated for each continuum step collapsed across target words. It was important to note that the reason that the proportion values, seen in Table 5.2, were not closer to either 0 or 1 was that targets were only lexical at one endpoint of the continuum (as seen in Figure 7.1): Place 1 was an /s/ phoneme, so 'Impressive' here was heard in correct form but for 'Condition' this continuum place corresponded to the greatest fricative variation. The opposite was true at place 15 which was a /ʃ/ phoneme, meaning the roles were reversed for the target words. The two intermediate points with the highest lexical bias were selected to be used in the main experiment: Continuum Place 5 and 11. The proportion of 'word' responses for both of the chosen intermediate points was 0.7.

To compare the results of the current pre-test to the results seen in Pitt and Szostak (2012) we had to change the Continuum from an absolute /s/ to /ʃ/ to 'word' to 'non-Word' by realigning Point 5 of both target words to the end of the continuum where that target word is presented is heard in its normal form. Table 5.3 shows the final 5-step continuum that will be used in the main experiment and the proportion of 'word' responses from the current Pre-test for each of the continuum points. The current pre-tests proportion values and pattern across the response range were similar to the 'diffuse' results seen for the late medial position (the closest condition to the current pre-test) in Pitt and Szostak (2012); although the pre-tested continuum points had lower values, aside from Point 5. Continuum point 1 had the largest difference with a reduced proportion of 'words' rating of 0.15, compared to around 0.45 in Pitt and Szostak (2012). This suggests that for the current targets, fricative variation at the 'non-word' continuum endpoint was more noticeable and consequently participants were less tolerant. They judged the targets containing this variation as lexical less than in the corresponding location in the Pitt and Szostak study (2012). The value of 0.15 observed for the current non-word end of the continuum (Point 1) was more closely matched to the word initial and word final

locations in the 'diffuse' condition in Pitt and Szostak. Both of these fricative variation locations showed a value below 0.2 for proportion of lexicality judgements at Point 1.

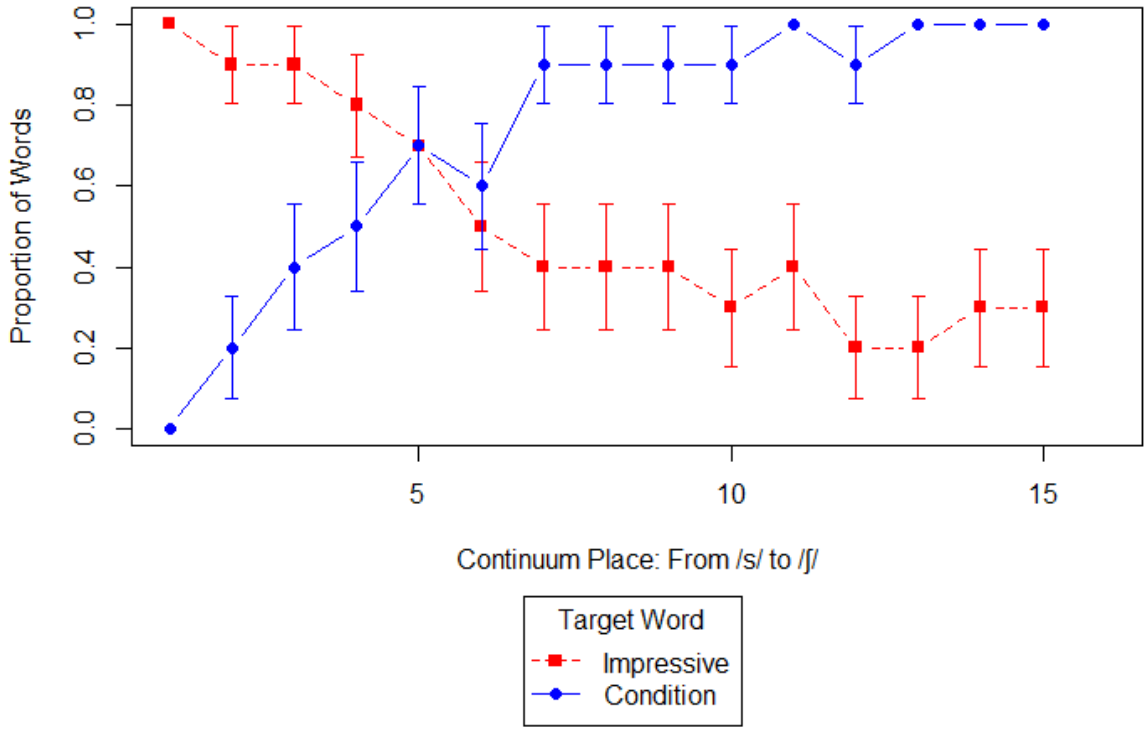


Figure 5.1- Pre-Test: The Proportion of Word Responses for each step along the continuum from /s/ to / ʃ / broken down by Target Word: ('Impressive' & 'Condition'). Continuum Fixed Points: Point 1- 100%/s/, 0% / ʃ/; Point 8- 50% /s/, 50 / ʃ/; Point 15- 0%/s/, 100% / ʃ/.

One of our original criteria was to have a range of response rates across the continuum, so we can measure the variation in effect of focused attention with stimuli that were across the lexicality range. The proportion values for the current Continuum covered a large percentage of the possible response range (0.15-1). This meant that the largest effects, seen Continuum medially, in Pitt and Szostak (2012) and to a lesser extent in Experiments 2 & 3 were unlikely to hit floor or ceiling effects for the middle points of the current Continuum as they did not fall at the extremes of the response range, either 0 or 1. Convergence towards the 'word' end of the continuum (Point 5 for our Continuum) was seen in Pitt and Szostak (2012) and we

would predict a similar phenomenon in the main experiment as listeners are highly likely to recognise the normal presentation of the target words and respond accordingly. This meant that the proportion value of 1 here should not result in a loss of effect due to it being a ceiling value. Additionally, we predict that focused attention would lower the 'word' proportion values and due to the pre-test being closest to an 'Unfocused' condition, in the main experiment an effect in this direction would still be viable.

Continuum Place	*Continuum Point*	*%/s/*	*% /ʃ/*	*Proportion of Word Responses*
1	1	100	0	0.5
2	2.1	95	5	0.55
3	2.2	90	10	0.65
4	2.3	85	15	0.65
5	2.4	80	20	0.7
6	2.5	75	25	0.55
7	2.6	70	30	0.65
8	3	50	50	0.65
9	4.1	5	95	0.65
10	4.2	10	90	0.6
11	4.3	15	85	0.7
12	4.4	20	80	0.55
13	4.5	25	75	0.6
14	4.6	30	70	0.65
15	5	0	100	0.65

Table 5.1- Pre-Test: The Proportion of Word Responses by Continuum Place, Continuum Point and the Percentage of /s/ and /ʃ/ for each Continuum Point. Highlighted are the Continuum Points that were used in the main experiment.

Continuum: Word to Non-Word	*Proportion of 'Word' Responses*
1	0.15
2	0.55
3	0.65
4	0.85
5	1

Table 5.2- Pre-Test: Final Continuum from 'Word' (Point 1) to Non-Word (Point 5) and the proportion of Word Responses.

5.4 Experiment 4: Speech Perception and Focus

In Experiment 4 we asked how focused listener attention affects a lexical decision task when there is pronunciation variation at a phonemic level. In the current study we updated the fricative variation to a word-medial location housed in new target words with a newly pre-tested continuum, described above, with the aim of testing whether participants' sensitivity to phoneme manipulation is reduced when attention is explicitly directed to the ambiguous phoneme.

5.4.1 Participants

A total of 40 students from the University of Edinburgh participated for a reward of £6.50 upon successful completion of the study. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties. Participants who had taken part in either the pre-test or Experiments 2 or 3 were excluded from taking part in the main experiment.

5.4.2 Design and Materials

Each trial was made up of a place holder sentence (e.g., '*She remembered to say…'*) followed by a target word that participants then had to make a lexicality judgement on. There were two types of target word: experimental targets and filler targets. The experimental targets were always tokens of the word pair: 'Impressive' and 'Condition'. The criteria for the selection of these target words is discussed above. Each experimental target was combined with each step of the continuum to create 5 variants. At one end of the continuum each experimental target was a word. The same target was made into a non-word at the other end of the continuum: 'Impressive' became 'Impreshive'; 'Condition' became 'Condision'. There were 12 Filler targets were made up of 6 matched word and non-word pairs. Each trisyllabic word had a matched equivalent that had a manipulated medial phoneme to make it into a non-word ('doubtingly' and its matched non-word equivalent 'doupingly').

All manipulated phonemes in the filler targets only differed in either place or manner of articulation. None of the filler targets contained an /s/ or /ʃ/ sound.

There were 10 place holder sentences included to increase the ecological validity of the task. They were short and context neutral, so that participants were not anticipating any certain entity. Each place holder could accommodate all experimental targets and filler targets. For example, "*She remembered to say*...". The experimental instructions gave the place holders context by stating that they related to a native British English speaker giving instructions about a word list. All sentences were between 15-21 characters long and finished with either "say" or "be". They followed the same structure so that the effect of the place holder sentence would be minimised.

A complete experiment consisted of 700 trials. There were 100 trials with experimental targets: A target with each continuum point (5 points) and presented with all 10 place holder sentences (50 trials). There were 2 target words (100 trials). Additionally, there were 600 filler trials: each of the 12 filler targets presented with each of the 10 place holder sentences (120 trials) and repeated 5 times (600 trials). The experimental targets made up 14% of trials with fillers presented for the remaining 86% of trials. This equalled the percentages of experimental targets (14%) and fillers (86%) seen in Experiment 2 in Pitt and Szostak (2012).

The experiment was broken down into 10 blocks of 70 trials. A block contained 10 trials with experimental targets. All 5 continuum steps contained in both target words presented with a different place holder sentence (10 trials). A combination of target word and continuum point was only ever presented with a specific sentence place holder once. For Example, 'Impressive' with Continuum point 1 was only every presented with any sentence place holder once. All 12 filler targets were presented with 5 different place holder sentences (60 trials). 30 of the trials occurred with a word filler target and 30 occurred with a non-word filler target. Each

combination of a filler target and a sentence only occurred once in a block and only occurred in 5 of the 10 blocks.

At 700 trials the experiment was very long and we were concerned about the viability of participants being able to concentrate long enough to complete the whole experiment. To assuage these concerns the experiment was split into 2 lists; each list always contained the 5 blocks. A participant only saw 1 List containing 5 blocks, so 350 trials in total. This meant that a participant only saw a target word variant with 5 of the 10 place holder sentences. They saw an equal amount of trials containing each target word (25 trials). They saw 300 filler target trials, half of which were word targets and half of which were non-word targets. They saw each combination of filler target and sentence place holder at least twice but were presented with some three times. The combinations that occurred 3 times in one list then occurred only twice in the other list.

The focused attention manipulation came in the form of the instructions that participants saw. There were two sets: (i) Focused and (ii) Unfocused. Participants only ever saw one set. The instructions accounted for the pronunciation variation as "mistakes". As noted above, the instructions included additional contextual information compared to Experiments 2 & 3. Apart from the inclusion of extra information, the instructions matched those seen in the previous experiments. The 'Focused' condition instructions alerted participants that possible mistakes would always be in the final word and that changes could be small and would be sound based and took place at the start of the final word. In the current experiment, we placed additional focus on /s/ and /ʃ/ phonemes by defining the task in greater details using these phonemes: "You may for example hear the speaker saying '*sh*' in the middle of a word when they mean '*s*', in which case they may have mistakenly produced a sound that isn't a word". In the 'Unfocused' condition the instructions simply stated that there could be some mistakes and did not emphasise which word within the sentence or the location within the word where these mistakes could

occur. The instructions here diverged from Pitt and Szostak (2012) because of the differing task demands. The inclusion of a place holder sentence meant that we had to identify the final word as being the target that participants had to make a lexical decision on. Both sets of instructions can be seen in full in Appendix A. The splitting of the experiment into 2 lists meant that participants (n=10) took one of 4 combinations of the experiment: List 1 with 'Focused' instructions; List 1 with 'Unfocused' instructions; List 2 with 'Focused' instructions and List 2 with 'Unfocused' instructions.

Comprehension questions were included after 20% of trials. These questions were added to measure participants' engagement throughout the task. These questions only followed trials which contained a word filler target because we could not ask participants to answer a questions about a non-word. The comprehension questions asked participants to select one of two choices: the target they just heard or a competitor word. The competitors were phonetically or semantically similar to the target word heard (e.g., "Confirmed" and the competitor target for this trial was "Conformed"). There were an equal number of comprehension questions (12) in each block.

All materials were produced by a native British English speaker. The speaker was instructed to produce the materials in a naturalistic manner. Sentence place holders and target items were recorded separately and repeated until a delivery approximating natural speech was achieved. The sentential contexts were always produced with the token "pen" as the final word, keeping the effects of co-articulation and prosody between the sentence and experimental target words constant throughout the experiment. The recording process for the target items is described above in the pre-test section. The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. All recordings were saved in a mono 48kHz .wav format.

5.4 3 Apparatus and Procedure

The visual and audio stimuli were presented using DmDX software (version 5; Forster & Forster, 2013) on a PC and a 15 inch monitor set at a 1024x768 resolution. Groups of up to 4 listeners were tested simultaneously across 2 rooms. The 2 computers in each room were separated by a divider meaning that participants could not see the computer screen of the other participant at any time during the experiment. If there were 2 participants in one session they were seated in different rooms. After reading an information sheet and filling in a consent form, listeners were seated at a computer and told to put on headphones that were attached to their computer. Although there could have been two listeners in a room simultaneously, there was unlikely to have been any noise distractions from the other participant due to the over-ear design of the headphone, which minimised ambient noise.

Participants then read through the practice trial instructions presented onscreen. These instructions matched the instructions of the main experiment and the distinction between 'Focused' and 'Unfocused' conditions was present in the practice trial instructions. The instructions asked participants to judge the final word of the sentence they heard as either a word or not by pressing a key corresponding to their choice. Quick and accurate responding was also stressed. Following this they performed 4 practice trials, which did not vary across participants. These practice trials comprised of 4 filler trials taken from the main experiment and were designed as a familiarisation phase.

Trials started with a count down marker of "###", "##", "#"after which the trial would begin. This countdown marker was used so that the participant's attention would be cued to focus on the auditory stimuli from the beginning. Combined with the start of the place holder sentence, "+++++" was displayed on the screen as a visual cue for the duration of the place holder sentence and target word. After this, participants had to select whether they thought the target was a word or not by pressing either the left or right 'CTRL' key. The 'Word' and 'Non-Word' responses

were written on the side of screen relating to the key that needed to be pressed to select that answer. The position of the 'Word' response matched the dominant hand of the listener, as self-identified at the beginning of the study on the consent form. For example, if the participant was right handed then it would appear on the right hand side of the screen. This meant that the 'Non-Word' response would appear on the side of the listener's weaker hand. Once either a selection had been made or the trial timed out after 3000ms, the next trial began automatically. After completing the practice phase, participants got the chance to ask any further questions of the experimenter. At this point the experimenter reiterated that participants should go with their initial response as to whether the stimulus was a word or not and that there was no wrong answers.

Participants then viewed the same instructions as during the practice trials again with the addition of an extra screen that alerted participant to the fact that comprehension questions followed some trials. Participants were told that for the comprehension questions they would have to choose from one of two answers that would be presented on screen. They should press the 'CTRL' key corresponding to the side of the screen the answer they wished to select appeared on. If the answer was on the right side of the screen then they should press the right 'CTRL' key. They then moved to the main experiment. The trials here followed an identical structure to the practice trials described above. However, a comprehension question followed 20% of trials. After a selection was made in either a normal or comprehension question trial the next trial began automatically. Participants got a break between blocks with them having to press a key to resume the experiment and move to the next block. The study took between 45-60 minutes to complete depending on the length of breaks taken by a participant.

5.4.4 Measures

The measures that were used were the proportion of word responses for each continuum point and the percentage of comprehension question that were answered correctly.

5.4.5 Analyses

We analysed participants' lexicality judgements. Our primary focus was the proportion of lexical responses and how this looked across the continuum when broken down by Focus and/or Target condition. All analyses were only on the experimental target data; filler trials were removed. We excluded trials where participants did not make any selection and the trial timed out. This accounted for 1% of all trials. For the purposes of analysis, we created a new Continuum variable: the Continuum factor discussed below is not the absolute 5-step continuum from /s/ to /ʃ/ above but a new 5-step continuum from non-word at point 1 to word at point 5. This was created by reversing the original /s/ to /ʃ/ continuum for the 'Condition' data, so that the /ʃ/ and, hence, the word end of the continuum was realigned with point 1. This meant that for the new Continuum predictor that both experimental targets were 'words' at the same point, Point 5. For each trial, if the participant selected a word we coded this as 1 and if a non-word then this was coded as a 0. Due to our dependent variable being binomial (whether a participant judged a target as a word or not), we employed the same analyses as in the previous studies (Experiments 2 & 3): a linear mixed-effects regression model with empirical logit transformed proportion data. This model was 'maximally specified' with both random intercepts and slopes, as well as their correlations varying by participants, as suggested by Barr, Levy, Scheepers, and Tily (2013). The reasoning for the choice of an empirical logit transformation was that we expected that at the Continuum endpoints there would be a lot of either a lot of 0s but few 1s, or vice versa. When this occurs logistic regressions tend to have problems converging. This problem is minimised when an empirical logit transformation is employed. The predictors we

used in the analyses were Focus (Focused and Unfocused), List (List 1 or List 2) and Half (1 or 2) which were between participants and Target ('Impressive' and 'Condition') which was within participants.

The comprehension question data was used as a check throughout the experiment: If a participant was consistently answering comprehension questions wrong, above 20%, then we would question the validity of their data and remove their remaining data. Each comprehension question was coded as 1 for a correct answer and 0 for an incorrect answer and then we created a percentage for each participant based on the number of correct responses.

5.5 Results

We first present the comprehension question results, as this affected which lexicality judgement data was taken forward to be analysed. The results of the lexicality judgement analyses were presented following this. All lexicality judgments were analysed in R (R Development Core Team, 2014) using the lme4 package (Version 0.999999-0, Bates, Maechler & Bolker, 2014). The p values were calculated using the lmerTest package (Version 1.2-0, Kuznetsova, Brockhoff & Bojesen, 2013).

5.5.1 Comprehension Questions

As described above, we wanted to check each participants' answers to the comprehension questions to decide whether the rest of their data should be included in the main analyses. We excluded trials where the comprehension question had timed out and not been answered. This accounted for 5% of all trials. The lowest comprehension question score was 94% of comprehension questions answered correctly. This was a sufficiently high percentage for all participants' data to include in the main analyses.

5.5.2 Proportion of Lexical Responses

The current study aimed to investigate the effect of 'Focused' attention on the proportion of word responses made by participants. Figure 5.2 shows that the proportion of word responses increased along the continuum from the non-word end (Point 1) to the word end (Point 5). However, there appeared to be minimal differences between Focus conditions with an average difference of 0.04 proportions of word responses. The Focused condition reduced participants' lexicality judgements: The largest variation (0.13) was seen at the midpoint of the Continuum (Point 3) with minimal variation at Points 2 (0.04) and 4 (0.02). There was convergence at the endpoints of the Continuum (Points 1 and 5): a minute difference in proportions of lexical judgements of 0.01 at Point 1 and identical proportions of 'word' responses for Point 5.

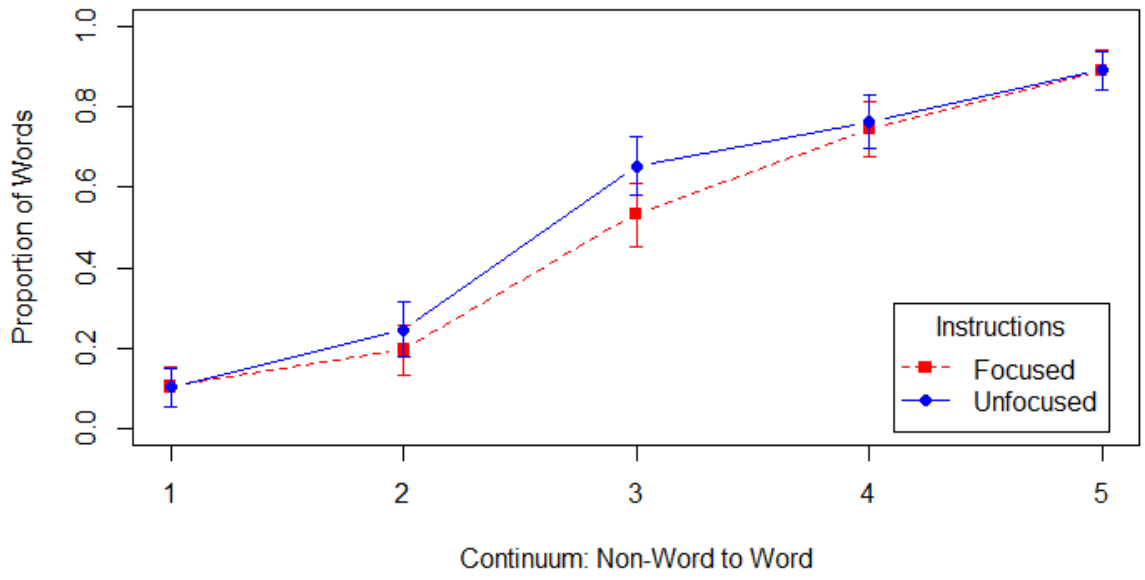


Figure 5.2- The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused).

We used empirical logit transformed proportion data and a linear mixed model with Focus and Target as sum coded predictors and this confirmed there was no effect of

Focus (t <1). The results seen here follow Pitt and Szostak (2012) in so much as that where differences exist between the focus conditions, the 'Focused' condition made participants less tolerant of the fricative variation but there is no measurable difference between the conditions. The lexicality judgement proportions are all lower than seen for the late-medial position in experiment 1 of Pitt and Szostak (2012).

In previous experiments there were large variations in the proportion of 'word' responses when broken down by Target. Figure 5.3 revealed a recurrence of this pattern for the current study. There were large differences between target words from Continuum point 3 onwards, with an average variance of 0.40 in proportions of 'word' responses. The variance between the targets decreased towards Point 5. The largest variation was again seen Continuum medially, with a decrease of 0.61 in participants rating of 'Impressive' as a 'word' (0.29) compared to 'Condition' (0.90). The /ʃ/ target, 'Condition', showed a strong lexical bias from Point 3 to 5 with values rising from 0.9-0.98 for proportion of lexical responses.

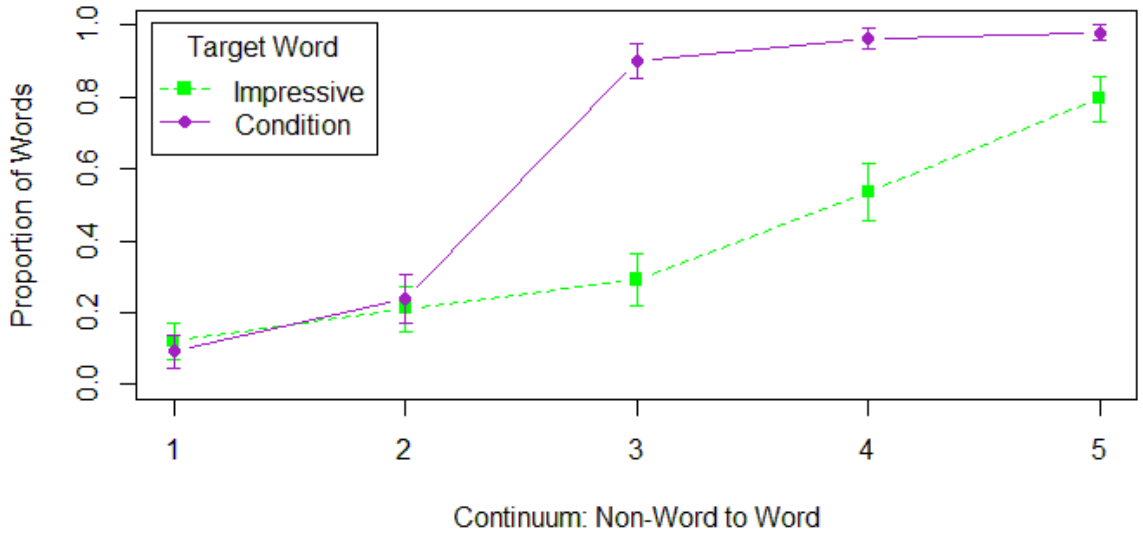


Figure 5.3- The proportion of word responses for each Continuum Point (Non-Word to Word) by Target word (Impressive and Condition).

There were minimal differences between targets seen at Continuum Points 1 and 2. Replicating experiments 2 and 3. The /s/ target word variant, 'Impressive', showed decreased word response for lexical judgements than the /ʃ/ target, 'Condition' at

every Continuum point, aside from Point 1. The consistency and size of the difference was reflected in Target producing a main effect in the linear mixed model described above (β= -0.58, SE=0.11, t=5.31, p <0.001). In comparison to the results seen in the previous experiments there were localised differences, such as the decreased values for Continuum Point 1 and 2 for 'Condition' but generally the pattern was consistent.

Figure 5.4 shows the Target conditions further broken down by Focus condition. The largest variation when broken down Focus conditions occurred continuum medially, seen in Figure 5.4. Encouragingly, this difference between Focus conditions was still observed when broken down by Target; focused attention reduced the proportion of 'word' responses in both target words: 0.1 for 'Impressive' and 0.15 for 'Condition'.

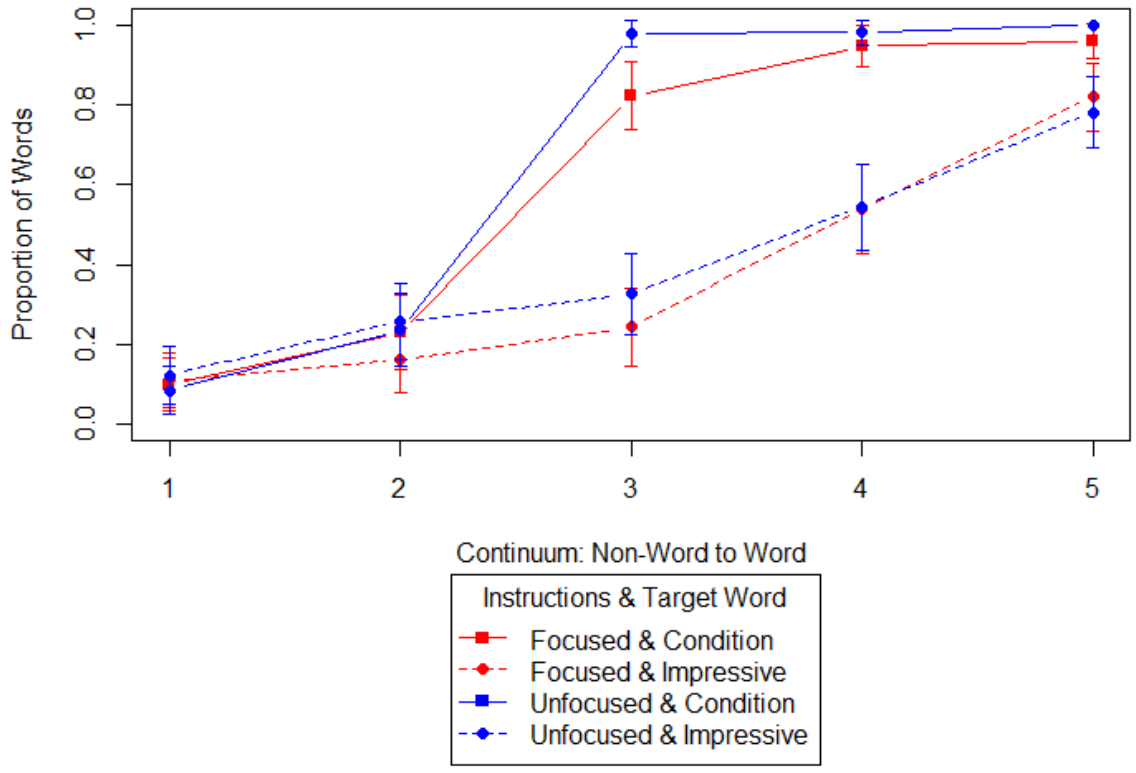


Figure 5.4-The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused) and then by Target word (Impressive and Condition).

There was minimal variation seen across the whole continuum: an average difference in proportions of 'word' responses between Focus condition by target words of 0.06 for the /s/ target and 0.05 for the /ʃ/ target. The only notable variance (above 0.1) between Focus conditions occurred Continuum medially: the variance at both Point 3s above and a 0.1 at Point 2 for 'Impressive'. The remaining differences in proportions of words responses by Focus condition were all less than 0.05 in both targets. Additionally, both target words showed convergence between Focus conditions at the continuum endpoints.

Breaking the target words down by Focus condition did not alter the pattern of differences seen when the responses were shown by Target alone; there were still large reductions in lexicality judgements for 'Impressive' compared to 'Condition' seen from Point 3 onwards. There was no reliable interaction effect between Target and Focus in our linear mixed model ($t<1$).

Due to the length of the complete experiment being 700 trials, the experiment was split into 2 lists as discussed above in the Design & Materials section. We wanted to check whether this was having an effect on how participants made lexicality judgements. We ran a separate linear effects regression model still with empirical logit transformed proportions with 'List' as a predictor and no effect was seen ($t <1$). The duration of the experiment was long (50-60 minutes) and repetitive with participants seeing 350 trials. Therefore, we wanted to check whether participants were paying less attention towards the end of experiment and whether this could possibly mask an effect. To guard against this, we created a new predictor, 'Half' which coded whether a trial occurred in the first or second half of the experiment. Running a separate linear mixed effects regression model still with empirical logit transformed proportions with 'Half' as a predictor revealed no effect ($t <1$).

Consistently for the results of the current study the largest differences between predictors were located at the midpoint of the continuum. At Point 3, the Continuum was truly ambiguous as the phoneme sound was constructed of exactly half /s/ and half /ʃ/, meaning the continuum point was midway between the two phonemes.

Therefore, if focused attention was to cause participants to be less tolerant of fricative variation and counter a lexical bias effect, then differences should be clear at this point. The lack of variance seen towards the endpoints of the continuum could have been counteracting an effect seen at the midpoints of the Continuum. Therefore, we ran the analyses again but only with responses for Continuum point 3. A linear mixed model with empirical logit transformed proportion data with Focus and Target as predictors was used to analyse this data. Starting with the effect of 'Focused' attention on word responses at this continuum point revealed a difference between Focus conditions (β= -0.31, SE=0.15, t=2.06, p= 0.046). Focused attention led to a reduction in the proportion of lexicality judgements of participants. For Target, there had been a large effect seen across the whole continuum, the biggest gap between proportions of 'word' responses for target words occurred at Continuum Point 3. Unsurprisingly, again there was a large effect of Target at the midpoint (β= -1.41, SE=0.16, t=8.50, p <0.001). In the analyses for all continuum data above there was no interaction between Focus and Target and that was repeated for only the Point 3 data, (t <1). Half and List both had no effect when analysed in the same way as described above but ran on only the Point 3 data (t's <1.5).

5.6 Discussion

In the current study we updated the target words used and moved the location of the phoneme variation to a word medial location to maximise the chances of observing our predicted attentional effect, as seen in Pitt and Szostak (2012). Unfortunately, as with the results of the previous two studies there was no reliable difference shown between the focused and unfocused attention conditions for responses across all continuum points. However the results here showed that there was a reduction in the proportion of word responses for the focused attention condition at the midpoint, continuum point 3. This difference remained when the responses were broken down by both Focus and target word. Both target words

showed a reduction in word responses at the midpoint, which differed to the pattern of results seen for the Focus conditions in experiments 2 and 3 where no differences were observed continuum medially for the /s/ target. This resulted in a reliable difference for responses between attentional conditions at continuum point 3.

Although the results were closer to our predictions than in the previous experiment they still failed to replicate an attentional effect across the continuum. As suggested in the previous studies, it is possible that due to the target words following sentence placeholders, participants had to comprehend preceding material that may have resulted in a reduction in their attention to the sentence final word when compared to the Pitt and Szostak study. The results also showed that the largest differences always occurred continuum medially and these differences were reduced towards the continuum endpoints by both Focus and Target conditions. This suggests that participants' sensitivity was affected most by the attentional manipulation when the phoneme variation was at its most ambiguous. In comparison to the differences seen at the endpoints of the continuum in Pitt and Szostak the current pattern of results is again reduced (2012).

Although in the current study we had updated the target words and the continuum location within the word, notable variance between target words across the continuum remained. As in the two previous studies the largest variation occurred at continuum point 3, although with a reduction in the gap for the current results. Again the /ʃ/ target 'Condition' received a significantly higher proportion of word responses in comparison to the /s/ target word 'Impressive', an effect which was especially prevalent in the middle points of the continuum. A similar pattern of results was observed in the previous experiment, where participants also judged pronunciation variation as less 'word' like when it was housed in the /s/ predicted target. This lends further weight to the proposal that participants may have been more comfortable with variation in the /ʃ/ target word over the /s/ target.

We had removed any obvious weaknesses in the selection of the current target words and they matched those which had driven results in the Pitt and Szostak

paradigm (2012). It follows that the lack of result is harder to justify aside from factors that we cannot investigate from the current results such as the acoustic nature of stimuli from Pitt and Szostak's or differences between British and American English or sensitivity to variation that may diverge between groups of participants from these respective countries. If our continuum was more natural sounding than Pitt and Szostak's then this could have influenced participants' sensitivity to the phoneme variation heard. Additionally employing two target words leaves the results susceptible to any variance that may be associated with individual target words.

Taken together the results of Experiment 4 did not wholly match our predictions. However, based on the reliable effect seen at the continuum midpoint and the lack of variance seen in the previous studies, the current paradigm represented the optimum methodology and materials for us to move to the next stage with the addition of disfluency to investigate the attentional account of disfluency processing.

CHAPTER 6

Experiment 5

6.1 Introduction

In Experiment 4, we found encouraging results with consistent variation between focused and unfocused conditions found at the continuum midpoint. Although, this variation was not reliably observed across the whole 5-step Continuum when we focused in on the midpoint an effect of focused attention was found. This was a partial replication of the focus effect seen in Pitt and Szostak (2012): In their study, the effect generalised across the whole continuum. There were a number of possible reasons for why we still failed to see an effect that extended across the whole continuum. The word medial target word pair demonstrated a large difference between /s/ and /ʃ/ target words, as seen for the word initial targets, and this remained a robust effect. The differences between the two target words were particularly prominent continuum medially. This pattern of differences between target words has been consistently observed in all speech perception studies that we have undertaken.

The word medial /ʃ/ target, 'Condition,' demonstrated a similar proportion of word responses from the midpoint of the continuum onwards to the 'word' end (Point 5) when compared to the pattern seen for these points in the results of both experiments across word locations for fricative variation in Pitt and Szostak. However at the 'Non-word' end of the continuum (Points 1 & 2) the values were markedly decreased to the corresponding values for a late medial word location seen in experiment 1 of Pitt and Szostak (2012). A late medial position was not included in Pitt and Szostak (2012) Experiment 2 but for their word medial location in this study there was similarly increased values to what was observed in experiment 4.

Despite the lack of a focused attention effect across the continuum, the progress of more consistent variation and a reliable effect at Point 3 meant that we thought that

the addition of disfluency would be useful in an exploratory capacity and could still speak to our predictions about attentional accounts of disfluency accounts. We aimed to investigate whether following a disfluency there was heightened attention by employing a speech perception paradigm. If disfluency were to drive increased attention to the incoming speech stream then this would be expected to impact upon low-level speech perception, resulting in increased sensitivity to incoming speech stimuli, as shown for the attention based perceptual effects seen in Pitt and Szostak (2012). Previous disfluency processing studies have shown similar effects: Hearing a disfluency may direct listeners' attentional focus, thus, causing them to more closely attend to the incoming acoustic stimuli (e.g., Collard et al., 2008). Therefore, the primary focus of the current study was to explore the additional effect that disfluency may have on the paradigm that was reported above in Experiment 4. The disfluency employed for the current paradigm was the filled pauses (i.e.,'uh' and 'um'). Although we made no specific predictions about how each type of filled pause might differentially affect lexicality judgements, we were interested to observe any differences following the different filled pauses.

Disfluency was introduced to the paradigm in the sentence place holders. It always occurred in the same sentential position: pre-target. In this position, disfluency is syntactically close to the target word and if disfluency was directing listener's attentional focus then the target word is the next speech heard. This maximised the chances of an effect, as there is no intervening phonemic information before the target word.

Focused attention was also manipulated in the form of the instructions. In the 'Focused' condition participants were informed as to the word and the location within a word where the phonemic variation would occur. This was not the case in the 'Unfocused' condition where listeners were just told that variation could be present: The instructions did not guide participants' attention to the final word or the word initial phoneme in this condition.

All participants heard disfluent productions, therefore if disfluency was shown to modulate attention then we would predict that this effect would be seen most clearly in the 'Unfocused' condition, as this condition served as the attentional baseline condition in the previous study. In this unfocused condition the disfluency would be predicted to create an attentional peak comparable to the 'Focused' instructions which would result in a similar decrease in proportions of 'word' responses. We detailed the rationale for this in the chapter overview in Experiment 2. We still predict that there should be a lexically biased attention effect between the 'Focused' and 'Unfocused' instruction conditions, similar to that seen in Experiment 4 and in Pitt and Szostak (2012). We make no predictions for those participants that see the 'Focused' instructions with the addition of disfluency, as we are uncertain as to how the two will interact.

Taken together our predictions centre on participants being less accepting of the manipulated pronunciations as 'words' following hearing a disfluent sentence place holder. Additionally we would predict a reduction in lexicality judgements for those participants that also receive the 'focused' instructions compared to those that saw the 'unfocused' instructions. If effects for either 'Focused' attention or disfluency were to be found, we would be interested to see if they extended across the continuum, how an effect breaks down by target word and the size of the effect relative to Pitt and Szostak's (2012) demonstration of this attentional manipulation effect and how this could relate to our ultimate question of whether disfluency has a similar effect in this paradigm and the implications this has for the attentional account of disfluency processing.

6.2 Disfluency Creation

We aimed to create tokens of disfluency that closely matched the phenomenon occurring in natural speech and that could generalise to instances of disfluency employed in other studies. Disfluency always occurred in the same location throughout the experiment: immediately prior to the target sentence-final word. The

motivation for this location is outlined in the introduction above. In summary, we suggest that it maximises the chances of an effect stemming from the disfluency by having no intervening phonemic information before the target word. If there was additional phonemic material between the disfluency and target then any possible increased attentional resources and heightened acoustic sensitivity stemming from this increase in attention, could have been reduced or nullified before the target was heard. This would mean that there would likely be no differences in participants' lexicality judgements and no effect would be observed for the current paradigm.

We used two variants of a filled pause in the current experiment: 'uh' and 'um'. By using two filled pause variants we reduced the repetitiveness of the disfluency heard, thereby, increasing the potential for attention orienting effects. Differences have been proposed (detailed in the literature review above) between differing tokens of filled pauses (e.g., Fox Tree, 2001) and we wanted to explore whether these differing fillers affected lexicality judgements differently and whether this had implications for the paradigm. Although this complicates the experiment by adding an additional manipulation to the paradigm, which reduced the observations per cell, there were still 500 observations of each type of disfluency that allowed us to test the differences between each type of filled pause against the fluent condition and against each other. The average duration of the disfluencies in the current study was 514ms (SD: 117ms). There was a difference of around 200ms between the averages of each filler: 'uh' 415ms (SD: 33ms) and 'um' 614ms (SD: 93ms). These disfluency durations were slightly longer than the equivalents seen in Fox Tree (2001).

All of the sentence place holders had a matching disfluent version that was recorded separately. The disfluent version featured one of either 'uh' or 'um' only. The average duration of the fluent sentence place holders was 1657ms and the average duration of the matched disfluent sentence place holders was 2471ms: A difference of 814ms. This difference was longer than the fluent presentation of the sentences with the addition of the average disfluency. This discrepancy was likely down to natural variance in production and additional pauses that surround the disfluency.

However, the difference was relatively small: 300ms or 18% of the average fluent place holder duration. This relatively short increase in duration was unlikely to have affected participants' lexicality judgments in a reliable manner, as the extra time afforded them no advantage in the task because regardless of the length of sentence place holder, the targets which judgments were made on were the same length between fluent and disfluent productions.

The disfluent sentence holders were recorded in the same session and manner as the fluent sentence place holders. The disfluent sentential contexts were always produced with the token "pen" as the final referent, keeping the effects of co-articulation and prosody between the sentence and experimental target words constant throughout the experiment. Participants were instructed to insert the disfluency as naturally as possible. The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. All materials were produced by a native British English speaker. The speaker was instructed to produce all materials in a naturalistic manner and the materials were repeated until this was achieved.

6.3 Experiment 5: Speech Perception and Focus

In the current study, we asked how 'Focused' listener attention and the inclusion of disfluency affected a lexical decision task when there was pronunciation variation at a phonemic level. The fricative variation remained at a word-medial location and we used the pre-tested Continuum and target words described in Experiment 4.

6.3.1 Participants

A total of 40 students from the University of Edinburgh participated for a reward of £6.50 upon successful completion of the study. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties.

Participants who had taken part in either the pre-tests or main study for Experiments 2, 3 or 4 were excluded from taking part in the current study.

6.3.2 Design and Materials

The materials and design followed the same structure as Experiment 4 but with the additional inclusion of disfluency. Each trial was made up of a place holder sentence followed by a target word that participants then had to make a lexicality judgement on. There were two types of target word: experimental targets and filler targets. Both sets of targets are described in detail in the design and materials section of Experiment 4 and are identical to stimuli used for that study.

The 10 fluent place holder sentences described in the previous design and materials section were reused again for the current experiment but with the inclusion of 10 matched disfluent place holders. A full description of the fluent place holders can be found in the design and materials section of Experiment 4. The disfluent place holders were identical to the fluent versions but with the inclusion of a filled pause at the end of the place holder. This means the disfluency would always be located immediately pre-target word. There were two possible filled pauses: 'uh' and 'um'. A disfluent place holder only ever occurred with one filled pause. Each filled pause variant occurred with half of the place holders: 5 disfluent place holders used 'uh' and 5 used 'um'.

A complete experiment consisted of 700 trials. There were 100 trials with experimental targets: A target with each continuum point (5 points) and presented with 10 of the 20 place holder sentences (50 trials). Each target occurred with 5 disfluent sentence place holders and 5 fluent place holders. Each specific target variant (a target word containing one step of the continuum) was presented with all possible place holders but only in one fluency condition: a place holder was never presented in both its fluent and disfluent condition for a single continuum point within a target word for each participant. There were 2 target words (100 trials).

There were 600 filler target trials: each of the 12 filler targets were presented with 10 out of the possible 20 place holder sentences (120 trials) and repeated 5 times (600 trials). Each filler target was presented with one variant from the 10 matched fluency pairs: 5 of the 10 place holders in each fluency condition. A filler target was never presented with both the fluent and disfluent version of the same sentence place holder. All possible combinations of fluency of sentence place holder (fluent and disfluent) with word and non-word filler targets combination occurred within in the experiment but not for each filler target. The combinations were: fluent place holder with a word filler target; fluent place holder with a non-word filler target; disfluent place holder with a word filler target and disfluent place holder with a non-word filler target. The proportion of trials for each of these combinations was equal at 25% of filler target trials each (150 trials). Therefore for the disfluent combinations, 75 trials occurred with the filled pause 'uh' and the remaining 75 trials were presented with 'um'. The experimental targets made up 14% of trials with fillers presented for the remaining 86% of trials. This equalled the percentages of experimental targets (14%) and fillers (86%) seen in Experiment 2 in Pitt and Szostak (2012). We did not present all combinations of sentence place holder with experimental and filler targets as this would have created 1400 trials, which would have required an unrealistic number of participants to be ran within the timescale of the current thesis.

The experiment was broken down into 10 blocks consisting of 70 trials. A block contained 10 trials with experimental targets; all 5 continuum steps contained in both target words presented with a different place holder sentence (10 trials). Either the fluent or disfluent version of each place holder from all 10 matched sentence place holders was used. Half of the place holders were from each fluency condition. A target word variant was only ever presented with a specific sentence place holder once. Each block contained a target word variant presented with a different sentence place holder alternating between a fluent or disfluent delivery. The remaining 60 trials in each block were filler target trials. All 12 filler targets were presented with 5 different place holder sentences (60 trials). 30 of the trials occurred with a ‘word’

filler target and 30 occurred with a ‘non-word’ filler target. Half of the trials were presented with fluent and half with disfluent sentence place holders. Each combination of a filler target and a sentence place holder only occurred once in a block and only occurred in 5 of the blocks. However, a ’Word’ filler target and its matched ‘Non-Word’ filler target could occur with the same target place holder within a block.

At 700 trials the experiment was very long and we were concerned about the participants’ ability to pay attention for the duration of the whole experiment. To assuage these concerns the experiment was split into 2 lists: each list contained a set 5 blocks. A participant only saw 1 List containing 5 blocks, so 350 trials. This meant that a participant only saw each target word variant (for example, ‘Impressive’ with continuum point 1) with 5 of the 10 place holder sentences (50 trials). They saw equal amounts of each target word (25 trials). They saw 300 filler target trials: 75 from each of the combinations described above. There were equal numbers of 'word' and 'non-word' filler targets (150 trials for each filler target type) and fluent and disfluent productions of sentence place holders (150 trials of each). Participants saw each combination of filler target and one of the fluency variations of a sentence place holder pair at least twice but were presented with some three times. The combinations that occurred 3 times in either list then occurred only twice in the other list.

The focused attention manipulation came in the form of the instructions that participants saw, as detailed in Experiment 4. The splitting of the experiment into 2 lists meant that participants took one of 4 combinations of the experiment: List 1 with Focused instructions; List 1 with Unfocused instructions; List 2 with Focused instructions and List 2 with Unfocused instructions. 10 participants completed each condition of the experiment. Comprehension questions were included after 20% of trials, matching Experiment 4; comprehension questions only followed fluent presentations with a 'word' filler target. The auditory stimuli matched those detailed

above in the Disfluency Creation section and described in Experiment 4. All recordings were saved in a mono 48kHz .wav format.

6.3.3 Apparatus and Procedure

The apparatus and procedure were identical to Experiment 4**.**

6.3.4 Measures

The measures that were used were the proportion of word responses for each continuum point and the percentage of comprehension question that were answered correctly.

6.3.5 Analyses

We analysed participants' lexicality judgements. Our primary interest was the proportion of lexical responses and how this looked across the continuum when broken down by Focus and disfluency conditions. All analyses were only on the experimental target data, filler targets were removed. We excluded trials where participants did not make any selection and the trial timed out. This accounted for 0.8% of all trials. For the purposes of analysis, we created a new continuum variable (Continuum): the continuum factor discussed below is not the absolute 5-step continuum from /s/ to /ʃ/ described above but is instead a new 5-step continuum from non-word at point 1 to word at point 5. This was created by reversing the original /s/ to /ʃ/ continuum for the 'Condition' data, so that the /ʃ/ point and, hence, the word end of the continuum was realigned with point 1. This meant that a continuum with 'word' end of the continuum for both experimental targets was created.

For each trial, if the participant selected a 'word' response we coded this as 1 and if a 'non-word' then this response was coded as a 0. Due to our dependent variable being binomial (whether a participant judged a target as a word or not), we employed the same analyses as in the previous studies (Experiment 2-4): a linear mixed-effects regression model with empirical logit transformed proportion data. This model was

'maximally specified' with both random intercepts and slopes, as well as their correlations varying by participants, as suggested by Barr, Levy, Scheepers, and Tily (2013). The reasoning for the choice of an empirical logit transformation was that we expected that at the Continuum endpoints there would be either a lot of 0s but few 1s, or vice versa. When this occurs logistic regressions tend to have problems converging. This problem is minimised when an empirical logit transformation is employed. The predictors we use in the analyses were Focus (Focused and Unfocused), Exp (Experiment 4 and 5) and Half (the half which the trial fell in: 1 or 2) which were between participants and Fluency condition (Fluent and Disfluent), Continuum (as described above Points 1 (Non-Word) to 5(Word)) and Target (Impressive and Condition) which were within participants.

The comprehension question data was used as an attention check throughout the experiment: If a participant was consistently answering comprehension questions wrong then we would question the validity of their data. For each comprehension question we coded 1 for a correct answer and 0 for an incorrect answer. We then created a percentage of correct responses for each participant.

6.4 Results

We first present the comprehension question results, as this affected the data to be analysed for the main experiment. The results of this lexicality judgement analyses are presented following this. All lexicality judgments were analysed in R (Version 2.15, R Development Core Team, 2014) using the lme4 package (Version 0.999999-0, Bates, Maechler & Bolker, 2014), p values were calculated using the lmerTest package (Version 1.2-0, Kuznetsova, Brockhoff & Bojesen, 2013).

6.4.1 Comprehension Questions

As described above, we wanted to check participants' answers to the comprehension questions to test whether the rest of their data should be included in the analyses.

We excluded trials where the comprehension question had timed out and not been answered. This accounted for 5% of all trials. The lowest comprehension question score was 93% of comprehension questions answered correctly. On this basis all participants' data was included in the main analyses.

6.4.2 Proportion of Lexical Responses

The current study was designed to explore the effect of focused attention and disfluency on the proportion of ' word' responses made by participants. The analyses below are first presented with each of the main predictors (Focus, Fluency and Target) collapsed across the remaining predictors. Then the interactions between these predictors are analysed, before we finally focus in on further analysis on the consistent pattern of differences seen continuum medially.

6.4.3 'Focus' Analyses

Figure 6.1 shows the Focus conditions collapsed across Fluency and Target for the whole continuum. There was a small amount of difference between the Focus conditions for the current study:

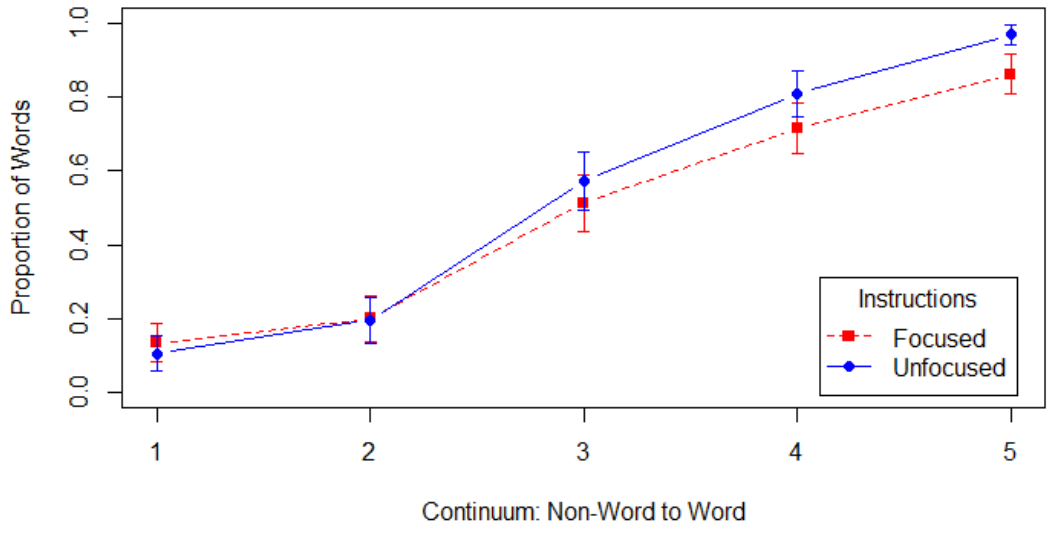


Figure 6.1- The proportion of 'word' responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused).

The 'Focused' instruction presentation had a reduced proportion of 'word' responses at every Continuum point but often there was only a minimal difference compared to the 'Unfocused' condition (average= 0.07). There was convergence between Focus conditions towards the 'non-word' end of the Continuum (Points 1 & 2) with close to matched proportions of 'word' responses seen (0.03 or below). From the midpoint of the Continuum the variation between Focus conditions increases, with the largest variation at Points 4 and 5 (-0.11 for the Focused condition).

We used empirical logit transformed proportion data and the linear mixed model described above with Focus, Disfluency and Target as sum coded predictors and this confirmed there was no effect of Focus, ($t < 1.4$). The results seen here followed Pitt and Szostak (2012) in so much as that where differences existed between the focus conditions the 'Focused' condition appears to make participants less tolerant of the fricative variation but there is no reliable difference between the conditions. The lexicality judgement proportions at each Continuum point all have lower values than for the equivalent point in the late medial position in experiment 1 of Pitt and Szostak (2012). There was no 'Focus' effect across the continuum in Experiment 4 and the differences observed here did not reach significance. However, in Experiment 4 a Focus effect was seen at Continuum Point 3 where the largest differences were seen. In the current study the greatest difference is not seen at the midpoint but at the 'word' end of the Continuum and this represented a change in trend from the previous study. The differences seen at Continuum 3 is analysed and discussed below. In Experiment 4 there were only fluent presentations and for the current study there was also fluent presentations. However, for the current study all participants heard both fluent and disfluent presentations. We wanted to test if a Focus effect was observed for these Fluent presentations alone. We excluded all disfluent presentations and ran the model again but excluding the disfluency predictor, as the data only included the Fluent variants of place holder sentences. There remained a lack of a Focus effect across the Continuum ($t < 1.1$).

The proportion of 'word' responses for both Focus conditions were relatively matched to Experiment 4: The 'Focused' condition here had an average difference of 0.03 to the equivalent in Experiment 4. The largest difference being a variation of 0.04 between Experiments at Points 1, 3 and 4; The 'Unfocused' showed a similar pattern with an average difference of 0.05 between Experiments and a maximum difference of 0.09 at the midpoint. These differences could have been in either direction (increase/decrease) between the Experiments. These figures serve to highlight the small differences between the Focus condition variants in each experiment suggesting that the inclusion of disfluency has not driven a change in the proportion of 'word' responses at any Continuum point when broken down by Focus.

We also wanted to clarify that statistically there were no differences between the proportion of 'word' responses by Focus condition between Experiment 4 and the current study. First, we created a new dataset collapsing across experiments. We then created a new predictor 'Exp': This coded which experiment a trial belonged to. Then we added this predictor to the model described above with the disfluency predictor excluded. We removed the disfluency predictor as there was no disfluency in the first experiment. There was no reliable difference in proportion of 'word' responses between Experiments ($t<1$). This model also supported our suggestion above that there were minimal differences between the proportions of 'word' responses between the Focus conditions and between experiments as there were also no interaction effects between Experiment and Focus ($t <1$).

6.4.4 Fluency Condition Analyses

Disfluency and the exploration of the effect it had on lexicality judgements was central to our predictions for the current study. This was the first paradigm to include a sentence holder with disfluency. The results collapsed across Focus and Target conditions can be seen in in Figure 6.2. The results showed that there was variation seen between Fluency conditions. There was an average difference of 0.10

in proportion of 'word' responses between the Fluency conditions. This figure does not take into account which direction the change was in.

The largest differences seen by Fluency were at Point 2 (0.18) and 3 (0.16) but interestingly in different directions at each point: At Point 2 the Disfluency condition shows a reduction in lexicality judgements compared to the Fluent presentation of place holder. However, 1 step further along the continuum at Point 3 and the Disfluency condition shows a sizeable increase over the Fluent condition. This equates to a large swing of 0.51 in proportions of 'word' responses for only 1-step in the Continuum in the disfluent condition.

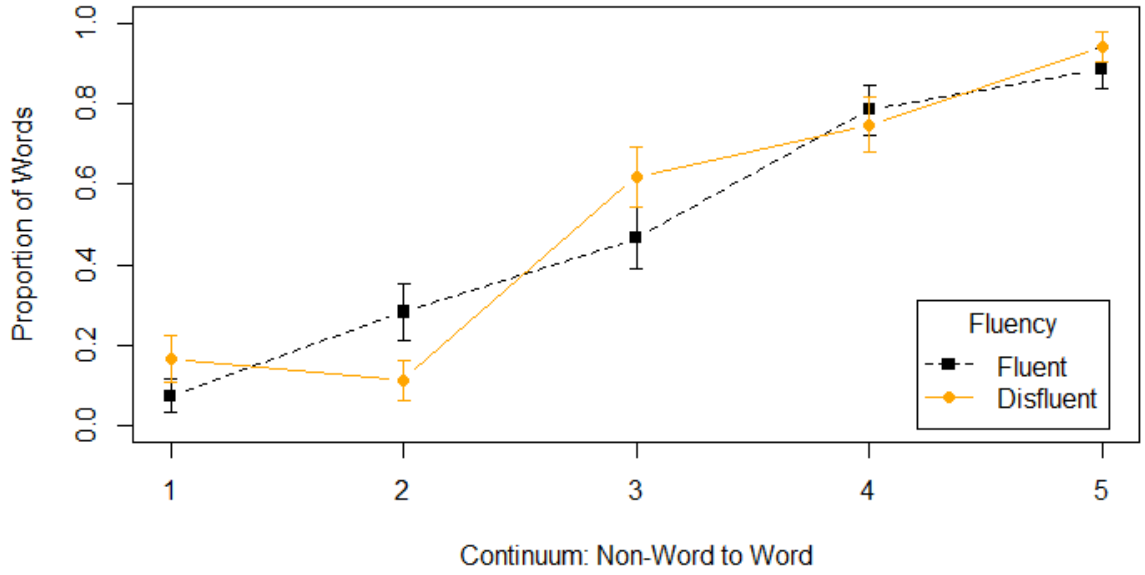


Figure 6.2- The proportion of word responses for each Continuum Point (Non-Word to Word) by Fluency Condition (Fluent and Disfluent)

At the endpoints and remaining medial position (Point 4) of the Continuum there were smaller differences between Fluency conditions (0.09 or less). However, despite there being differences across the whole continuum there was a lack of reliable effect for Fluency condition in the linear mixed model, ($t < 1$).

As discussed above, disfluency was realised in the current study using two different filled pauses: 'uh' and 'um'. Figure 6.3 shows the Fluent condition against the disfluency condition broken down by disfluent place holders presented with 'uh' and those presented with 'um'. Across the continuum there were only minimal differences between 'uh' and 'um' presentations (max 0.02 proportion of words for Points 1, 4 and 5) and they followed very similar patterns across the whole continuum. The largest differences between disfluency types were seen at Points 2 (0.04 proportion of words) and 3 (0.06 proportion of words). These points also represented the largest differences between Fluent and disfluent presentations. The 'um' presentation had the largest magnitude of change between these two points with a shift of 0.56 in proportion of 'word' responses, compared to 0.46 for 'uh'. This is a difference of over half of the scale between these two Continuum points.

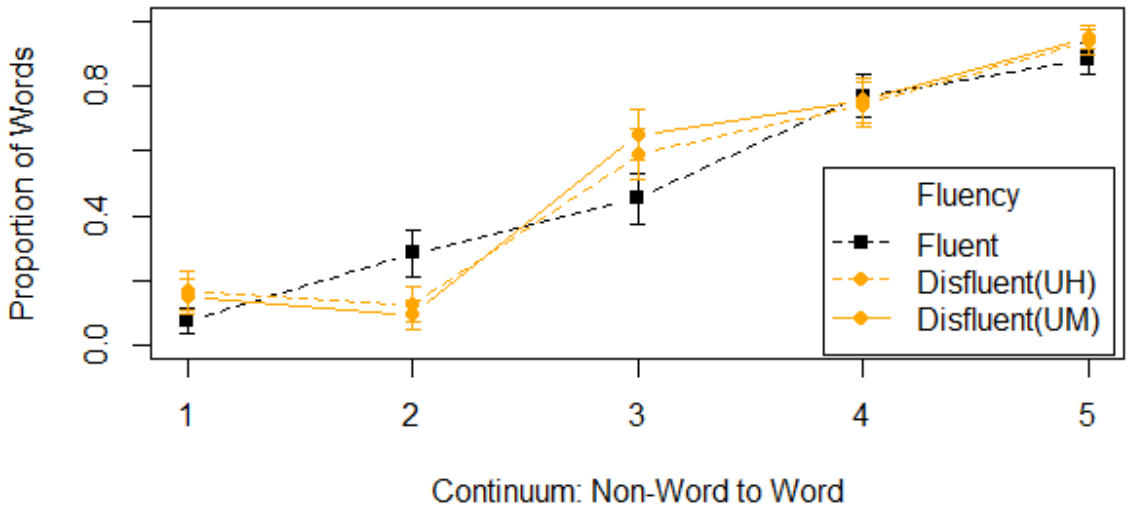


Figure 6.3 - The proportion of word responses for each Continuum Point (Non-Word to Word) by Fluency Condition (Fluent and Disfluent) with Disfluency broken down by Filled Pause ('UH' and 'UM').

The two variants of filled pauses may have driven participants to employ different behaviours or the same behaviours but in varying magnitudes when modelled, as

detailed in the introduction. By comparing the disfluency variants to one another and the single filled pause variants alone against the fluent condition we could see if they were reliably different. First, we subsetted only the disfluent presentations and ran a linear mixed model with empirical logit transformed proportions with the same predictors (Target, Focus, Continuum) as the model described above but exchanged the Disfluency predictor for a 'Disfluency type' predictor which labelled the filled pause presented for a trial differently for 'uh' versus 'um'. There were no effects or interactions for any combination of predictor for this model (all t's <1.) Next, we wanted to test each filled pause singularly against the fluent condition. We created 2 new datasets where 1 of the 2 filled pause variants was excluded, leaving only the data of one variant remaining: For the the 'uh' dataset, the presentations with 'um' included were removed and the 'uh' presentations removed from the 'um' dataset.

The 'uh' dataset followed the full data set with no Focus effect across the whole continuum ($t<1$). Target still had an effect but the size of the effect was reduced when compared to the complete data set ($\beta= -0.76$, SE=0.36, $t=2.09$, $p= 0.04$). The main focus of breaking the Disfluency predictor down into its constituent filled pauses was to explore to if the 'uh' and 'um' filled pauses caused a different result compared to the grouped disfluency predictor. For the 'uh' dataset described here, there was a reliable disfluency effect across the whole continuum ($\beta= -1.52$, SE=0.34, $t=4.47$, $p<0.001$). There were no other main effects or interactions shown for this disfluency variant.

Moving onto the 'um' dataset and there was similarity shown in the pattern of results to those observed in the 'uh' dataset. The manipulation of Focus did not have any effect ($t<1$). The different Targets created difference but again there was only a marginal effect compared to the complete data set ($\beta= -0.75$, SE=0.39, $t=1.93$, $p= 0.06$). There was a repetition of a main effect of disfluency across the Continuum for 'um' ($\beta= -1.38$, SE=0.37, $t=3.68$, $p<0.001$). There were no other main or interaction effects shown across the continuum.

6.4.5 Focus and Fluency Analyses

Continuing with the analysis of the impact disfluency in the current study, Figure 8.5 shows the Fluency conditions (Fluent/Disfluent) broken down by Focus. The relationship between these predictors was of interest as we wanted to see if there were any interactions, as this had implications for our predictions for the current study, as outlined in the introduction. Unfortunately, no interaction effect was seen in the linear mixed model ($t < 1$). However, the pattern seen for the Fluency conditions when collapsed across focus conditions is recurrent here; there is still a large increase in values for both 'Focused' and 'Unfocused' instruction types with a disfluent presentation between Continuum points 2 and 3. The fluent presentations for both Focus conditions show little variation.

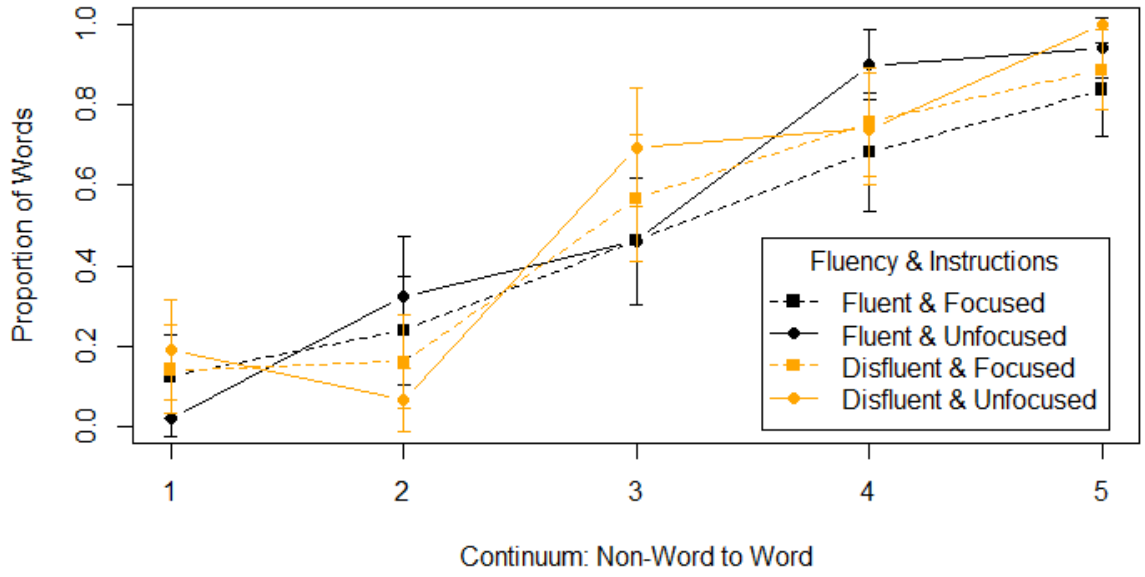


Figure 6.4 -The proportion of word responses for each Continuum Point (Non-Word to Word) by Fluency (Fluent and Disfluent) and then by Instruction Condition (Focused and Unfocused).

Comparing the differences between 'Focused' and 'Unfocused' instructions in the Disfluent condition revealed an average difference of 0.08 proportions of 'word' responses across the continuum. A comparable average was observed for the

variation between Focus conditions for the Fluent Condition: A difference in proportion of 'word' responses of 0.11. The largest difference between the Focus conditions in the Disfluent condition occurred Continuum medially with an increase of 0.15 for the 'Unfocused' instructions. This variation was interesting because when the Disfluent condition was broken down by Focus, it showed the same pattern observed for Disfluency alone (shown in Figure 8.3) with a large increase in values between Point 2 and 3. Additionally at the midpoint Disfluency was exerting a stronger influence in the 'Unfocused' instruction type, as noted above, 0.62 compared to 0.40 in the 'Focused' condition. There was no matching increase observed between the Focus conditions at Point 3 in the Fluent condition; the difference was only 0.01 proportions of 'word' responses. The largest difference for the Fluent condition when broken down by Focus of instructions was at Point 4: An increase of 0.22 proportions of 'word' responses for the Unfocused Condition over the Focused condition. This was in the direction that we would have predicted, with Focused attention making participants more critical of the ambiguous phoneme sound. However, there was only a small difference of 0.04 proportion of 'word' responses seen for this point when the responses were broken down by just Focus condition.

6.4.6 Target Analyses

The current study shows large differences in the proportion of 'words' when broken down by Target and collapsed across Focus and Fluency conditions: an average of 0.21 difference in proportions of 'word' responses between the targets across the continuum. Figure 8.2 reveals a recurrence of the robust pattern seen in Experiment 4 for Target. This was unsurprising due to the same materials being used here as in the previous experiment. The inclusion of disfluency seemed to have little effect on the proportion of responses or pattern of results for Target words across the Continuum.

In the previous studies (Experiments 2, 3 & 4) the /s/ target word variant always showed lower scores for lexical judgements when differences existed between the target words. 'Impressive' showed the same pattern of lower proportion values

compared to 'Condition' for the current study: The largest difference of 0.51 occurred at Continuum point 3. The differences between the target words extended to point 4 with 'Impressive' again having a reduction of 0.33 in 'word' responses compared to 'Condition at this Continuum point. We also noted the sharp increase in lexicality judgements between Point 2 and 3 for the /ʃ/ target for only 1 step along the Continuum: A jump of 0.66 in 'word' responses. There were minimal differences between Targets at Continuum Points 1, 2 and 5 (0.03 or less). The size of the differences seen Continuum medially were driving the robust Target effect observed in the linear mixed model, (β= -0.37, SE=0.13, t=2.77, p= 0.01).

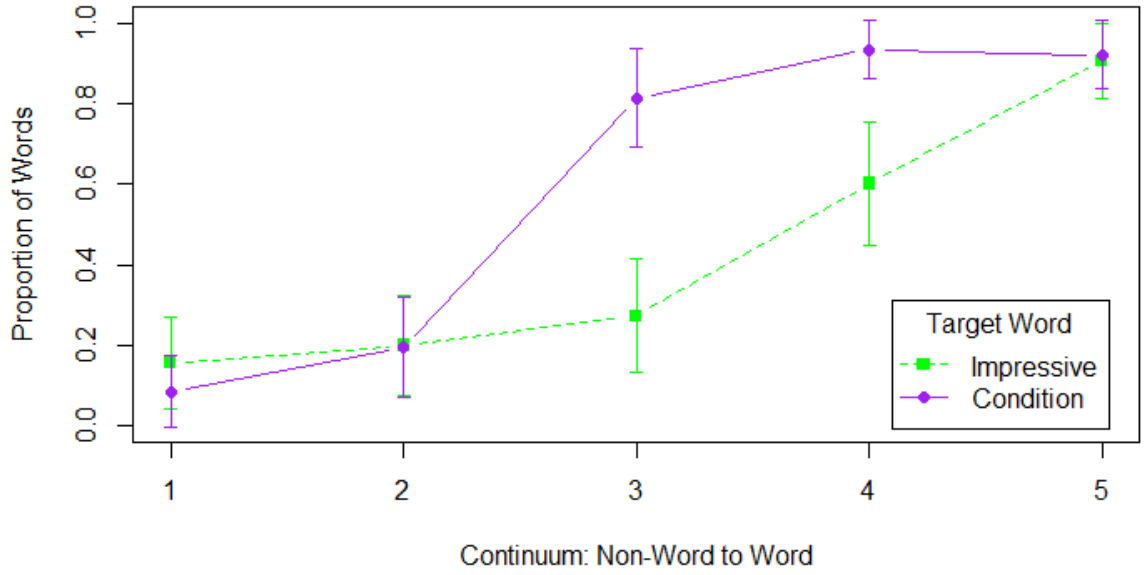


Figure 6.5- The proportion of word responses for each Continuum Point (Non-Word to Word) by Target word ('Impressive' and 'Condition').

In comparison to the results seen in the previous experiments (Experiment 2- Figure 5.3, Experiment 3- Figure 6.2) there were localised differences, but the general pattern of large Continuum medial variation between target words with convergence towards the end points of the Continuum was constant. Compared to the matched materials of Experiment 4 there was no notable differences, supported

by a lack of interaction effect seen between the 'Exp' variable and Target (t<1) when collapsed across experiments, as described above. This suggests that the inclusion of disfluency did not have any robust effect when broken down by Target. There was an average difference of 0.05 between proportions of 'word' responses for each Target between experiments.

6.4.7 Fluency and Target Analyses

The pattern of results for the target words when broken down by Fluency conditions follows a similar pattern to that seen in Figure 6.5 for Target alone. Variation is again seen Continuum medially when the Fluency conditions were broken down by Target word as seen in Figure 8.5.

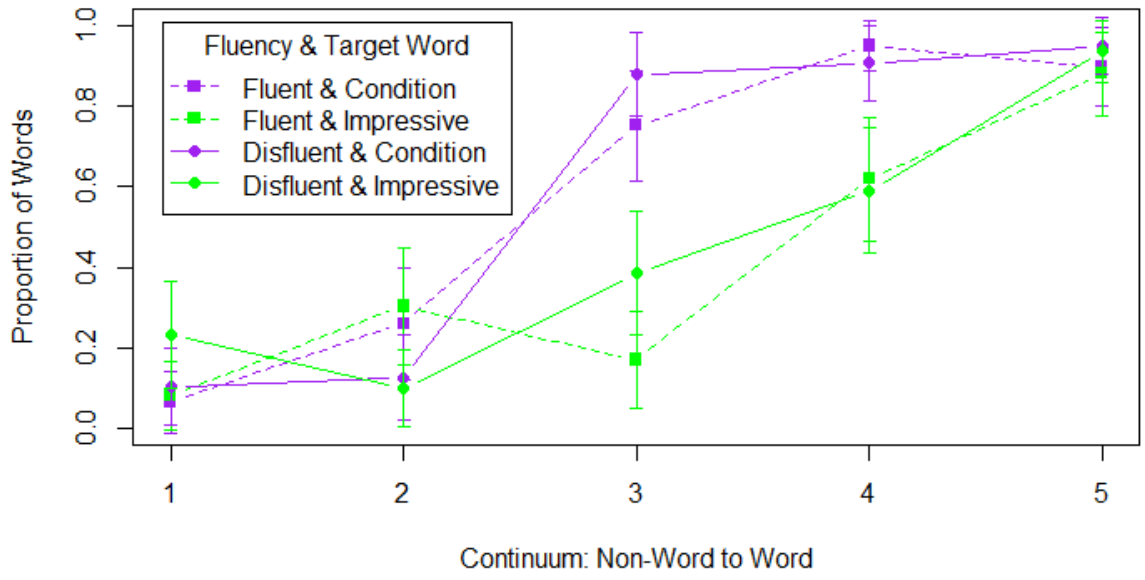


Figure 6.6-The proportion of word responses for each Continuum Point (Non-Word to Word) by Fluency (Fluent and Disfluent) and then by Target word ('Impressive' and 'Condition').

The pattern of a large increase in tolerance of fricative variation between Points 2 and 3 in all instances which have included the Disfluent condition can be observed again here for both Targets. The /s/ target 'Impressive' showed a reduction of 0.20 below the Fluent value of 'word' responses at Point 2; this rose to an increase of 0.19

at Point 3. A similar pattern was observed for 'Condition', with the Disfluent variant being 0.15 lower than the Fluent equivalent at Point 2 but with an 0.14 increase above at Point 3. This led a massive increase of 0.77 in proportions of 'word' responses for only a single Continuum step for the /ʃ/ target, 'Condition'. This value was 0.29 bigger than the same Continuum step seen for the Fluent presentations of 'Condition'. 'Impressive' saw a smaller increase of 0.26 between Point 2 and 3 in the Disfluent condition. This compared to a difference of 0.13 in the Fluent Condition for the same target between Points 2 and 3. There was limited variation between the Fluency conditions for the rest of the continuum: An average of 0.12 across the continuum for 'Impressive' and a slightly lower average of 0.08 for 'Condition'. As in the disfluency and Focus interaction above, there was no interaction effect between Fluency and Target seen in our linear mixed model, ($t < 1$).

6.4.8 Focus and Target Analyses

In Experiment 4 there was no interaction between Focus and Target.

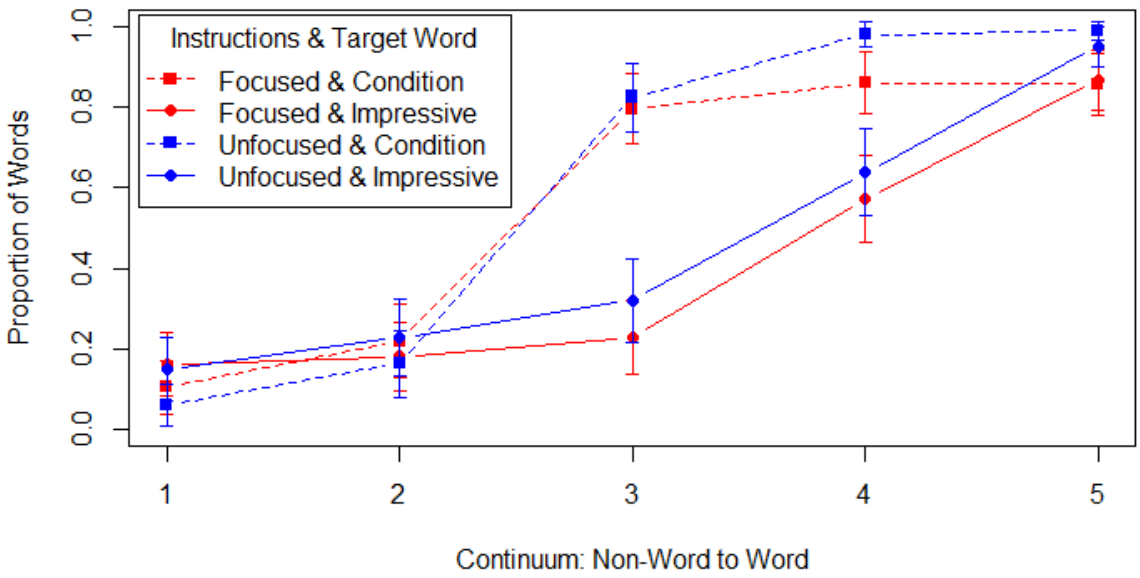


Figure 6.7- The proportion of word responses for each Continuum Point (Non-Word to Word) by Focus Condition of Instruction (Focused and Unfocused) and then by Target word ('Impressive' and 'Condition').

Figure 6.7 shows the results for the same predictors collapsed across Fluency conditions for the current study. The pattern of results for Targets when broken down by Focus Condition was similar to that seen in Figure 8.1 for Target alone: The largest variation between targets was seen Continuum medially with convergence of the Target words towards the endpoints of the continuum. 'Condition' had an increased proportion of 'word' responses for both Focused and Unfocused Conditions in the middle of the Continuum. The /s/ target, 'Impressive', and the /ʃ/ target, 'Condition' had equivalent average differences in proportion of 'word' responses between the Focused and Unfocused conditions across the continuum: A magnitude of 0.07 for 'Impressive' and 0.08 for 'Condition'. However, 'Impressive' showed a more consistent average reduction in the proportion of 'word' responses for the Focused condition across the continuum (-0.07) compared to 'Condition' (-0.04). The lack of interaction effect between these predictors shown in Experiment 4 was replicated in the linear mixed model built for the current study, ($t < 1$). The interaction between Target, Focus and Fluency was not graphed and there were no interaction effects seen ($t<1$).

6.4.9 Half Analyses

The current study was long (50-60 minutes) and repetitive with participants completing 350 trials. Therefore, we wanted to check whether participants were paying less attention towards the end of experiment and whether this could have impacted on the results seen above. To investigate this, we created a predictor, 'Half' which coded whether a trial occurred in the first or second half of the experiment. Running a separate linear mixed effects regression model still with empirical logit transformed proportions with 'Half' as a predictor revealed a marginal effect (β= -0.14, SE=0.08, t=1.77, p= 0.08). There was a reduction in the total proportion of 'word' Responses across the Continuum from 0.52 in the first Half to 0.49 in the Second. This may suggest a slight drop in attention. However, we propose this marginal effect was more likely due to increased sensitivity to the task as it progresses: Participants become more adept to the task.

A decrease between halves of the experiment was seen in Pitt and Szostak for responses at the word-final phoneme location in Experiment 1 (2012): There was a drop from 0.25 to 0.06 at the 'non-word' Continuum endpoint (Point 1) between the first and second half of the experiment for Participants in the 'diffuse' (Unfocused) Condition. This equates to a drop of 0.19 in proportion of 'word' responses. In comparison, the decrease for participants in the Unfocused condition for the current study for the same continuum point is from 0.11 to 0.10, a drop of 0.01. The largest difference in proportions of 'word' responses across the halves of the experiment is for Unfocused instruction at Continuum Point 3 with a reduction from 0.64 to 0.51: a drop of 0.13. A value still below that seen in the Pitt and Szostak study.

6.4.10 Analyses Performed on Continuum Point 2 and 3

The largest variation in the majority of the predictors analysed here came Continuum medially. This was a repetition of the pattern seen in Experiment 4. In this previous study, a Focus effect at the midpoint that did not exist across the whole Continuum led to the inclusion of disfluency in the current study. At the endpoints of the Continuum there was a tendency for convergence or a similar proportion between conditions.

Due to the repetition of differences seen Continuum medially, we again focused in on Continuum Point 3. At the midpoint, the Continuum was truly ambiguous as the phoneme sound here was half /s/ and half /ʃ/. Therefore, if focused attention or disfluency were to cause participants to change tolerance of fricative variation then this should occur at this point, as there was no preference for either of the /s/ or /ʃ/ phonemes. The lack of differences seen towards the endpoints of the continuum could have been counteracting effects at the midpoints of the Continuum. Therefore, we ran the analyses again but only with data from Continuum point 3. We employed the same linear mixed model with empirical logit transformed proportion data with Focus, Fluency and Target as predictors to analyse the Continuum point 3 data.

In Experiment 4, there was an effect for Focused attention ('Focused' instructions) at Continuum point 3 observed but this was not replicated for the current study (t <1.1). However, there was a repetition of the Target effect seen in all iterations of the current paradigm, (β=-1.24, SE=0.17, t=7.16, p <0.001), although Target did not interact with either Fluency or Focus (t's <1).

There were no effects observed across the continuum for the Fluency predictor collapsed across the filled pause variants. However, when analysed at Continuum point 3 a reliable effect was found between the Fluent and Disfluent conditions (β=0.38, SE=0.17, t=2.17, p= 0.036). As above for the whole continuum, Fluency and Focus did not interact at Point 3 (t <1.3). However, Target, Fluency and Focus showed a marginal interaction effect (β= -0.29, SE=0.15, t=1.95, p= 0.06).

The Disfluency included in the sentence place holders contained either an 'uh' or an 'um'. These had each been found to be significant across the whole continuum when the presentations with the remaining filled pause were removed. A Disfluency condition effect was again robust at the Continuum midpoint when analysed in the same way as described above, using only one of the filled pause variants, but ran only on the Point 3: 'uh' (β= -1.34, SE=0.53, t=2.54, p= 0.017); 'um' (β= -1.67, SE=0.47, t=3.53, p= 0.002). We also analysed how the List and Half predictors came out at the Continuum midpoint: List had no effect when analysed in the same way as described above but run only on the Point 3 data (t's <1). However, a main effect was shown for Half (β= -0.23, SE=0.09, t=2.52, p= 0.014).

A recurrent pattern in the results across the continuum was variation at Point 2 compared to Point 3, with the Disfluent condition consistently driving a reduction in the proportion of 'word' responses compared to the Fluent variant at Point 2 and then the opposite pattern one Continuum step along at point 3. This pattern was observed when disfluency was broken down by focus and target. We again ran the model on just the Continuum Point 2 data to see if there was any robust effects. At this point the only predictor that reached significance was Fluency condition (β= -0.4,

SE=0.14, t=2.82, p= 0.008). No other predictors or interactions showed reliable effects (t's <1.4).

A pattern of large shifts in proportion values between Continuum Point and 3 co-occurred with the disfluency predictor and continued to be observed when fluency was broken down by Focus and Target. With the confirmation of a disfluency effect at both Continuum point 2 and 3, we joined the two datasets to see if this effect held. The disfluency effect was in different directions but there was still difference away from the Fluent condition. There was no disfluency effect for the dataset made of Point 2 and 3 (t<1).

6.5 Discussion

In the current study we built on Experiment 4 with the inclusion of disfluency into half of the sentence place holders heard by all participants. Our central prediction was that, following on from hearing one of the disfluent sentence place holders, participants would show a decrease in the proportion of 'word' responses, similar to the attentional effect observed for the focused instructions condition at Continuum point 3 in Experiment 4 and across the continuum in Pitt and Szostak (2012). This was not the pattern of results seen: 'word' response proportions did not differ reliably by 'Fluency' across the continuum.

The pattern of results observed did show some differences between the fluency conditions, especially at Continuum points 2 and 3. At Point 2 the sentence holders containing disfluency matched our predictions and led to a reduction in the proportion of 'word' responses. However, at Point 3 the Disfluent condition created a difference in the opposite direction, it led to an increase in lexicality judgments at this point. This went against our predicted pattern of results. This equated to a large increase (0.51) in 'word' responses for this one step along the Continuum. This rise equated to over half of the available scale**.** Disfluency appeared to be causing two different behaviours for participants with this reversal seen after only a single Continuum step. This pattern of differences between Points 2 and 3 was consistently

seen when disfluency was included, even when the responses were further broken down; this pattern was seen when the responses were broken down by both fluency condition and focus condition and again for both fluency condition and target condition.

This repetition of the pattern of differences seen led us to again focus in our analysis on Continuum point 3 and additionally point 2 to explore if this difference equalled a reliable effect. At both of these points, when analysed separately, a reliable effect was observed between fluency conditions. However, this effect disappeared if the analysis included data from both Points 2 and 3.

The interaction between disfluency and the focused attention was of interest as we were unsure as to how these two predictors would impact upon participants' lexicality decisions, especially when participants heard a disfluent sentence place holder in the Focused attention instruction condition. There were no interaction effects seen between these two predictors. For Points 2 and 3 where the greatest variation had occurred between fluency conditions, when further broken down by instruction condition a similar pattern of results emerged, with the largest difference seen at these points. The difference for the Focused attention condition with the disfluent presentations was a reduced amount of variation away from the fluent condition, compared to those participants who had seen the unfocused instructions. It appeared that disfluency was negating the predicted effect of focused attention at these points. Although, when these predictors were analysed only on data at Points 2 and 3, no interaction effects were observed.

The disfluency condition being investigated here included two variants of filled pause: 'uh' and 'um'. In previous studies as detailed in the literature review, these filled pause variants have been proposed to elicit different behaviour from participants (e.g., Fox Tree, 2001). It follows that we wanted to investigate whether they were influencing participants' behaviour differently in the current study.

When both of the filled pause variants were analysed together but with a predictor that labelled whether it was an 'uh' or an 'um' heard by participants there was no reliable effect between fluency conditions. This suggests that there was a lack of difference between the impacts for each of the filled pause variants that were employed in the current study.

However, when the data was subsetted by filled pause variant, this created two new data sets, one that contained just the instances where participants heard 'uh' and the other with the remaining 'um' instances. These singular filled pause data sets were then analysed for differences between fluency condition and a robust effect was found for both. We did not make any specific predictions for the current study based on the filled pause that participants heard, as this was not a central research question for this series of experiments. However, it appears that including more than one variant of filled pause can impact the reliability of the disfluency predictor being testing, as a disfluency effect was observed when only analysed with a single filled pause variant.

We suggest that although the pattern of results for both types of filled pause was the same across the continuum, an explanation for the increased reliability of a disfluency effect when broken down by filled pause variant was because there were duration differences between the filled pauses. There has been evidence that a delay of any kind can facilitate linguistic processing (e.g., Bailey & Ferreira, 2003; Corley & Hartsuiker, 2011). The 'um' variants were an average of 199ms longer in duration that the 'uh' variants meaning that this allowed participants an increased amount of processing time which led to the 'um' disfluencies exhibiting the slightly more pronounced pattern observed, compared to the 'uh' filled pauses. Then when the filled pause variants were tested individually this led to less difference in the disfluency condition than when the two variants were included in the same data set, hence, driving a more reliable effect. In future studies, removing this source of variation in the disfluent condition would strengthen the paradigm.

In Experiment 4 there was no reliable difference shown between the focused and unfocused attention conditions for responses across all continuum points. There is no attentional effect observed between instruction conditions again for the current study. The differences that did occur between the Focus conditions was in the direction predicted, with the Focused instruction condition showing a reduction in the proportion of 'word' responses compared to the participants who saw the unfocused instructions. There was no interaction seen when both Focus conditions and Target were used to break down the responses. In Experiment 4, there was an effect for Focused attention ('Focused' instructions) at Continuum point 3 but this was not observed for the current study ($t < 1.1$). This was an interesting observation because the only paradigmatic difference between Experiment 4 and the current experiment was the inclusion of disfluency. As noted above in the discussion of the interaction of Fluency and Focus, the addition of disfluency to the paradigm appeared to be counteracting the attentional effects observed in Experiment 4. This is discussed further below.

The same target words that were used in the previous experiment were employed in the current study. There was a repetition of the Target effect seen in the previous studies. Again the /ʃ/ target 'Condition' received a significantly higher proportion of word responses in comparison to the /s/ target word 'Impressive', especially prevalent in the middle points of the continuum, with convergence towards the Continuum endpoints. This lends further weight to the proposal that participants may have been more comfortable with variation in the /ʃ/ target word over the /s/ target. This pattern of differences between target words remained when the responses were further broken down by Focus and Fluency conditions.

As noted in the previous study, the target words used in the current study did not contain the target word weaknesses, such as being a compound noun, that were identified in the earlier speech perception studies and matched those which had driven results in the Pitt and Szostakl paradigm (2012). It follows that the large differences between targets is harder to justify aside from factors that we cannot

investigate from the current results such as the acoustic nature of stimuli from Pitt and Szostak's study or differences between British and American English or sensitivity to variation that may vary between groups of participants from these respective countries. If our continuum was more natural sounding than Pitt and Szostak's then this could have influenced participants' sensitivity to the phoneme variation heard. Additionally just employing two target words leaves the results susceptible to any variance that may be associated with individual target words.

A marginal effect was recorded between the halves of the experiment, with a slight drop in the proportion of word responses for the second half. We propose that instead of representing a drop in attention it instead was more likely due to increased sensitivity to the task as it progresses: Participants become more adept to the task. Either way, the effect was marginal and was not impacting on the results observed in any reliable way.

Taken together these results suggest that disfluency was having a variable impact upon participants' proportions of lexicality judgments depending on the continuum point, with the most consistent differences and reliable effects between fluency conditions found at Continuum points 2 and 3. We had predicted that if there was heightened attention following a disfluency, as proposed in the attentional account of disfluency processing, that participants would be less accepting of the fricative variation, resulting in a reduction in the proportion of targets classified as 'words' in comparison to the fluent presentations. This finding would follow the focused attention effect seen in the experiment 4. However, this was not observed. This expected pattern of decreased 'word' responses was seen at Point 2 but a step along the Continuum at Point 3, disfluency was causing a reliable increase in lexicality judgments compared to the Fluent condition.

We propose that an explanation for this varied impact of disfluency was that there could be variable processing depending on the strength of lexical bias of the phoneme variation heard by the participants, which could lead to the change in effect seen between points 2 and 3. In the previous study the focused attention

instructions had driven participants to be more sensitive towards the phoneme variation heard, whereas, following a disfluency participants need not be more aware of the fine acoustic detail to successfully complete the task of making a lexicality judgement about the target heard. This could create divergent task demands between the focused attention condition and the disfluency condition that may result in varying levels of attention to the signal. Task demands crossed with attention have previously been shown to impact upon lexicality effects (e.g., Cutler et al., 1987). Similarly, Miller et al. (1984) reported that changing the task demands so that listeners' focus was directed to only the target word and away from the surrounding sentence context caused the disappearance of a context effect. In the current study, the focused attention condition did direct participants' attention towards a target word in comparison to the unfocused condition. This suggests, in line with our predictions, that listeners may have been changing their strategy towards the task between the focus conditions, leading to a greater reliance on bottom-up processing.

However, encountering disfluency introduced a different attentional effect that altered the task demands in both Focus conditions. As seen above, when the responses were broken down by both Fluency and Focus condition, following disfluent productions and the focused instructions co-occurring there was a reduction in the difference away from the fluent conditions, compared to participants who heard disfluent productions following unfocused instructions. This suggests that without the influence of the instructions drawing focused attention to the fine acoustic detail within a target, the increased attention afforded by disfluency acted as a facilitator effect that made participants less sensitive to the phoneme variation. Therefore, when disfluency and focused attention co-occurred they caused the participants to focus in again on the fine acoustic detail and the effect of the disfluency was less pronounced. However, in the unfocused condition the disfluency effects observed were larger, as there was less emphasis on the phonetic detail being heard and more on being decisive in the task of creating lexicality judgements.

An explanation for the variance in lexical bias between continuum points would be a cognitive load effect. If after encountering disfluency there was an increase in attention, this could lead to an increase in cognitive load experienced by the participant. Under cognitive load participants' have been shown to rely more on lexical influences (Mattys & Wiget, 2011). It follows that at Point 3 where the phoneme sound was truly ambiguous that the strength of lexical bias would be greater than at point 2 as the phoneme heard is closer to expected phoneme, which could explain the variable impact. This would provide support for a variable attentional mechanism as proposed by Mirman, McClelland, Holt and Magnuson (2008) who suggest that it be added to interactive models of speech perception to account for the variable perceptual processes that occur due to either implicit task demands or explicit focused attention. The possible benefit of variable phoneme classification following disfluency would be quicker and more efficient comprehension that is likely to aid in successful message transfer. In short, task demands were responsible for variable attending that was creating a notable cut-off between the level of lexical bias experienced at Point 2 and Point 3 resulting in the large swing in proportion of 'word' responses seen between these two consecutive steps.

In terms of the attentional account of disfluency that the current study set out to investigate, the disfluency effect seen here may be compatible in terms of an attentional modulation. However, this account also proposes that predictive or top-down processing ceases following disfluency. The disfluency effect observed at the Continuum midpoint worked counter to the way we predicted, due to the varying task demands for the fluency conditions when compared to the focus conditions. The disfluency appeared to be driving heightened attention to incoming signal but with the addition of varying lexical level activation dependant on the phoneme variation heard. The results here were in no way conclusive.

Overall, the results here are mixed, with a lack of either disfluency or focused attention effect seen across the current continuum. However, we have seen that

disfluency did have an impact continuum medially, with the results suggesting that disfluency and the role of attention are modulated by task demands and the cognitive load experienced by the user. In relation to our original aim of supporting the attentional account of disfluency, there were no definitive answers. The effects seen here are only based on responses to two target words, meaning that it is difficult to generalise these findings as being truly representative of reality. Investigating the phenomenon on more target words would explore whether individual target word variance is impacting any of the effects seen here and this is our aim with the next study.

CHAPTER 7

Experiment 6

7.1 Introduction

The last experiment in this thesis combines elements from the past four studies into an updated and improved paradigm that allows us to thoroughly investigate the impact of disfluency on the word-medial phoneme variation, with a view to exploring whether this supports an attentional account of disfluency processing, as detailed in the literature review.

We had previously proposed that if encountering disfluency led to heightened attention then it may impact processing in the same way as the explicit instructions employed in the Pitt and Szostak paradigm (2012). However, we were unable to produce a replication of an attentional effect across the Continuum using our sentence-based paradigm and an instructional attention manipulation. In all experiments the largest consistent differences between predictors has been seen Continuum medially. This led us to focus in our analyses on the midpoint in the previous two experiments. Experiment 4 revealed a Focused attention effect at the Continuum midpoint. Experiment 5 revealed a disfluency effect at the same point but no repetition of the effect of Focus seen in the previous experiment. In the previous experiment there was also a distinctive pattern seen with the inclusion of disfluency between points 2 and 3: a large fluctuation from a reduction in 'word' responses at Point 2 compared to the fluent condition to an increase above the fluent condition at Point 3. We had predicted that disfluency would lead to a decrease in proportion of ‘word’ responses across the Continuum but the pattern of results observed did not match this. The Continuum medial variance that was seen in Experiment 5 is focused on in the current study.

Although we found effects in Experiments 4 and 5, large variations between the target words were seen Continuum medially. This pattern of variance between /s/ and /ʃ/ targets has been seen in all of the previous speech perception experiments across a number of sets of target words, a robust effect of target has been seen in all of them. However, due to the relatively low number of target words tested these targets could have been anomalous and not representative of variance between typical /s/ and /ʃ/ words. In the previous experiments, we had opted for single pairs of target words so as to exert greater experimental control. However, this came at the risk of being able to generalise across beyond the target words being tested.

Experiment 6 was a partial replication of Experiment 5, which adapted the previous paradigm in the following ways: an increased number of target words; inclusion of only a single variant of continuum medial variation; a single filled pause variant 'um' and new filler targets. Firstly, we increased the number of target words: 16 each of /s/ and /ʃ/ targets. This meant that effects would have to generalise across a number of target words. These target words were presented using a single medial Continuum point, either point 2, 3 or 4. In the previous studies this was where we had seen the largest variance and the fricative variation at these points represented the maximal chance for the predictors to affect participants' lexical decisions. For each target word we chose the continuum point that resulted in the proportion of 'word' responses which was closest to 50% during our pre-test. As observed in all previous experiments, there were large differences between target words continuum medially: the/s/ target word consistently was observed with a decreased value of word responses Continuum medially compared to the /ʃ/. In the current paradigm, our method of continuum point selection should result in both of the target words should having roughly equivalent proportions of 'word' responses. Therefore, we can see the effect of the predictors between target words when the target words have the same proportion of 'word' responses. Essentially, we reordered the phoneme variation by proportion of word responses rather than continuum point.

The filler targets were also updated. In the previous studies none of the filler targets contained an /s/ or /ʃ/ sound, whereas, for the current study they all contained either an /s/ or /ʃ/ phoneme. The 'word' to 'non-word' manipulation involved exchanging the normal phoneme for the fricative sound at the other end of the continuum: /s/ phonemes and /ʃ/ were swapped. The motivation for this change stemmed from that in the previous studies the experimental targets' fricative manipulation differed to the phonemes heard in the filler targets. In the current paradigm, all manipulations were based on same phoneme sounds: the experimental targets and filler targets could not be distinguished from the change of phoneme sounds being manipulated.

Targets remained in sentence place holders and the instructions again contained the Focused attention manipulation. We also kept the word medial location for fricative variation in the current study due to the larger variance seen for this word position compared to the studies using the earlier world initial location. The disfluent presentations varied slightly from the Experiment 5: only the filled pause, 'um' was employed for the current study. We chose to employ only one because our primary focus was on investigating the attentional account of disfluency processing as a whole, not testing between specific filled pauses. Although each filled pause has been suggested to interact with language comprehension differently (e.g., Fox-Tree, 2001) there were no robust differences seen between 'uh' and 'um' in the previous study and including both could add more variance for no theoretical gain. The filled pause variant 'um' appeared to show give rise to slightly larger variation in Experiment 5, although this was not a reliable difference. This was why it was chosen over 'uh' for the current study. The pre-target location of the disfluency was repeated in the current study, due to the success seen with this disfluency location in Experiment 5.

Following the results observed in Experiment 5 we have updated a number of our predictions for the current study. Firstly, the prediction for the focused attention manipulation provided by the different sets of instructions remains constant from the previous experiment; after the 'focused' instructions participants would be less

likely to respond to a target with a 'word' response than in the 'unfocused' condition. Hence, the instructions would drive an attentional modulation that would increase a participants' sensitivity to bottom-up processing of the phonemic variation.

The prediction for the impact of disfluency for the current study was based on participants hearing ambiguous phoneme variation, as at Continuum point 3 in the previous experiment. Therefore, following the results observed in Experiment 5, we predict that participants would rate targets as 'words' *more* often when preceded by a disfluency when compared to a fluent condition. This means that disfluency would be making participants more accommodating of the phoneme variation heard.

These two predictions propose that the Focus and Disfluency conditions act in different directions. In Experiment 5, when these two conditions overlapped, the incidence of disfluency lead to an increase in the proportion of 'word' response values seen continuum medially compared to the fluent conditions. This impact of disfluency was reduced when disfluency occurred with the Focused instructions compared to the unfocused variant. However, neither of these differences resulted in a reliable effect.

Therefore, we predict the same pattern for the current study, when the instruction and fluency conditions co-occur this would lead to an increase in participants rating targets as a 'word' for both of the disfluent conditions compared to both of the fluent conditions. Although, we predict the magnitude of effect between the fluency conditions to decrease when focused attention co-occurs with the disfluent condition.

The predictions for those conditions which contain disfluency run counter to the predictions made in the previous speech perception studies in so much as they predict that the incidence of disfluency will instead increase the proportion of 'word' responses. If these predictions are supported then it would lead us to question the attentional account of disfluency in its current form. These results would not necessarily rule out a heightening of attention after encountering disfluency, instead

it suggests that the role of attention may be variable and based on the task demands and cognitive load that the listener is experiencing.

7.2 Continuum Creation

The creating of the continuum is of paramount importance as it contains the fricative variation that the rest of study is predicated on. We employed points 2, 3 and 4 from the continuum pre-tested and used in the previous experiments (4 & 5). The continuum creation details are described in 5.2.

7.2.1 Target Word Selection

Due to the increased number of targets that the new paradigm required we selected new target words. We opted for 32 target words: 16 /s/ and 16 /ʃ/ targets. The word-medial location for the fricative variation was retained. All 32 targets were 2 syllable words, matched for length (Average of 7 characters for both /s/ and /ʃ/ targets) and containing only one incidence of either /s/ or /ʃ/ phonemes to accommodate the fricative variation. The average duration for /s/ targets including the fricative variation was 633ms (SD: 87ms) and for /ʃ/ was 609ms (SD: 66ms): a difference of only 24ms. In the previous study we had used 3 trisyllabic targets but due to the number of targets required we moved to 2 syllable words, so there was consistency across all targets. These targets also met the basic criteria that at one end of the fricative continuum a 'word' was created and at the other end a 'non-word', for example, /s/ variant- 'decide' and 'deshide'; /ʃ/ variant- 'machine' and 'masine'.

The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. All materials were produced by the same native British English speaker as for all previous speech perception experiments (2, 3, 4 & 5). We again kept the speaker constant so there was consistency across experiments and that effects could not be attributed to a change in speaker. The speaker was instructed to produce the materials in a naturalistic manner. The target items were recorded using neutral sentence place holders to minimise list effects on the

pronunciation of the target and keep the production replicating natural spoken language as much as possible. This context was then excised to leave just the target. The method and process of recording stimuli was also repeated from previous experiments, so as to make the experiments as comparable as possible in terms of auditory stimuli. All recordings were saved in a mono 48kHz .wav format. A complete list of the targets can be found in Table 7.1.

7.3 Pre-Test 1

Our current paradigm hinged on being able to create a cohort of target words that all produced a relatively equivalent proportion rate of 0.5 'word' responses using one of the medial continuum points. The pre-test placed each of the three possible continuum points (2, 3 & 4) in each of the target words and then tested participants' responses to each target word to measure which target was closest to our goal value of 0.5 proportion of 'word' responses.

7.3.1 Participants

A total of 24 students from the University of Edinburgh participated for a reward of £4 upon successful completion of the study. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties. Participants who had taken part in either the pre-tests or main study for Experiments 4 or 5 were excluded from the current study.

7.3.2 Design and Materials

Each of the 160 trials was constructed of a single word target that participants had to judge either as a 'word' or 'Non-word'. There were two types of target: Experimental items and Fillers.

Each of the 32 Experimental targets were heard once by participants: 16 /s/ targets and 16 /ʃ/ targets. Each target was presented with only a single variant of the three continuum points being tested: Continuum medial points 2, 3 and 4. This created 3 variants of each target word that were each presented in 3 separate conditions, a

single participant only ever heard one. Each condition presented all 32 experimental targets with at least 10 targets containing each continuum point being tested. Each condition contained an additional trial of two of the 3 continuum points. However, across the experiment the number of trials containing each continuum point housed within each target were balanced. The occurrence of both /s/ and /ʃ/ target words with each of the continuum points was also balanced.

There were 128 filler targets which participants heard: 64 matched and 'word' and 'non-word' target word pairings. Each two syllable filler 'word' had a matched equivalent that had a manipulated medial phoneme to make it into a 'non-word', for example, a 'word' filler was 'cassette' and its matched 'non-word' equivalent was 'cashette'. To turn a 'word' into a 'non-word', the fricative phoneme was exchanged for the fricative phoneme at the opposite end of the continuum for normal usage, meaning that /s/ and /ʃ/ were swapped. The filler targets were identical across all 3 conditions. All items were randomly presented in a single block. Experimental targets made up 20% of trials, this value was slightly higher than the 17% seen in Experiments 4 and 5.

The instructions for the current study were equivalent to the 'unfocused' condition in previous experiments because they did not draw attention to location within the word where the fricative variation occurred. However, direct comparisons cannot be drawn because of the different task demands. In previous studies, participants were responding to a target word occurring at the end of a sentence place holder, whereas, for the current study listeners were making lexical decisions on single words.

The auditory stimuli recording details were as described above for the target words. The filler target items were recorded using neutral sentence place holders to minimise list effects on the pronunciation of the target and keep the production replicating natural spoken language as much as possible. This context was then excised to leave just the filler target. The method and process of recording stimuli was also repeated from previous experiments, to make the experiments as

comparable as possible in terms of auditory stimuli. The filler targets were recorded in the same session as the sentence place holders (described below) and the experimental targets detailed above.

7.3.3 Apparatus and Procedure

The visual and audio stimuli were presented using DmDX software (version 5; Forster & Forster, 2013) on a PC and a 15 inch monitor set at a 1024x768 resolution.

Up to four listeners could be tested simultaneously across two labs. Each lab housed 2 computers that were separated by a divider meaning that participants could not see the computer screen of the other participant at any time during the experiment.

After reading an information sheet and filling in a consent form, listeners were seated at a computer and told to put on headphones that were attached to their computer. Although there could have been two participants in a room simultaneously, there was unlikely to have been any noise distractions from the other participant due to the over-ear design of the headphone, which minimised ambient noise. Participants then read through the instructions presented onscreen. The instructions asked participants to judge the speech heard as either a word or not by pressing a key. Participants then moved to the data collection phase of the experiment. The structure of the trials matched the previous speech perception experiments (2, 3, 4 & 5), apart from upon the display of "+++++" it was the target word that began playing instead of a sentence place holder. Participants still had to select whether they thought the target was a word or not by pressing either the left or right 'CTRL' key. The time out value was reduced to 1800ms to recognise the shorter duration of the single word stimuli being presented. The study took approximately 15 minutes to complete. The procedure used for the current study matched the pre-test for Experiment 4.

7.3.4 Analyses

We analysed participant's proportion of 'word' responses for each of the medial continuum points within each target word. All analyses were only run on the

experimental target data, filler targets were removed. We excluded trials where participants did not make any selection and the trial timed out. This accounted for 4% of all trials. For each trial, if the participant selected a word we coded this as 1 and if a non-word then this was coded as a 0. To avoid confusion, the continuum discussed below is an absolute continuum ranging from a majority of /s/ fricative phoneme at Point 2, an equal split of 50% of each /s/ and /ʃ/ phoneme at Point 3 and a majority of /ʃ/ fricative phoneme at Point 4.

7.3.5 Results

The primary task of the pre-test was to select the variant from the three medial continuum points (2, 3 & 4) that was closest to the 0.5 proportion of 'word' responses value we required for each target word. Figure 7.1 shows the values of each of the 3 continuum points for each of the 32 target words containing. Notably, this graph shows that there is large variation by Target; this suggests that the effect of fricative variation was dependent on the word context it was contained in. However, for the vast majority of target words the expected pattern of 'word' responses by Continuum point is observed. The proportion of 'word' responses for each target hinges on the expected phoneme; the lowest proportion of 'word' responses was seen when the fricative variation of a continuum point was furthest from this expected phoneme and the highest proportion of 'word' responses was seen at the point of least fricative variation away from the expected target phoneme. At the midpoint, Continuum point 3, we expected an intermediate value that reflected the balanced nature of the phonemes in the fricative variation at this point. So for /s/ targets, continuum point 4 contained a majority of /ʃ/, so we expected this to produce the lowest values for targets containing the /s/ phoneme and this was observed in the current study. This preference by listeners was expected as the increased lexical bias at this point meant they are more likely to label a stimuli as a 'word' more often when the fricative variation heard in the target is closer to the fricative sound heard in a 'normal' production. This pattern can be observed in Figure 7.4 where the data is collapsed across targets but broken down by phoneme and continuum point. There

was notable variance between the proportion of 'word' responses at the outer continuum points (2 & 4) by target phoneme (/s/ & /ʃ/).

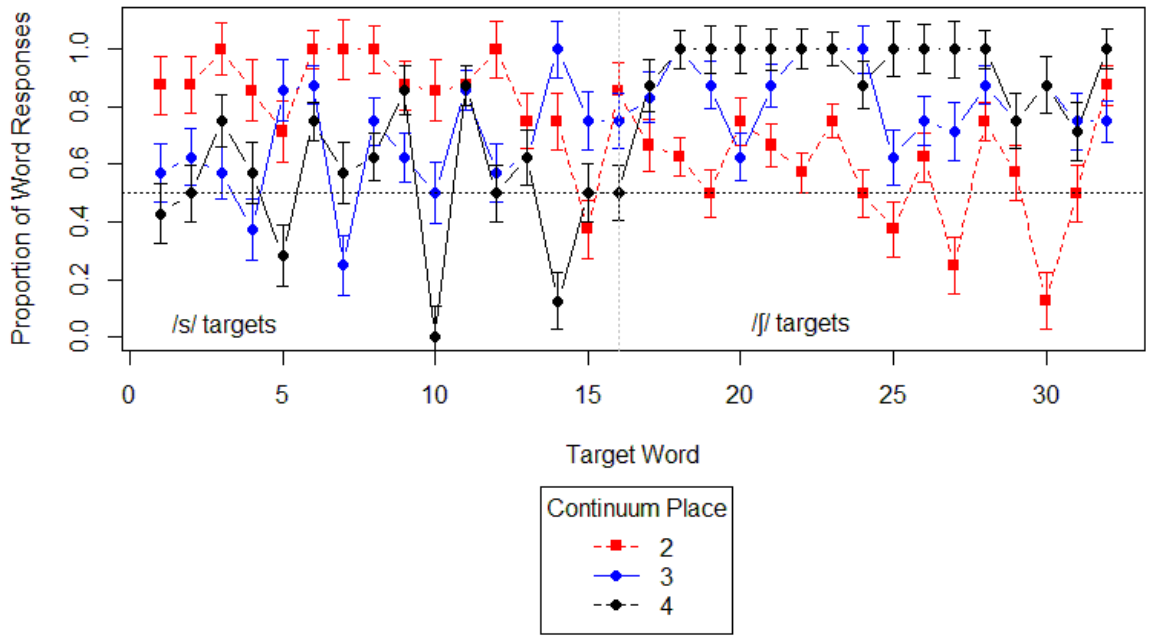


Figure 7.1- The Proportion of 'word' responses for each of the 32 target items broken down by the variant containing continuum Points 2, 3 and 4. Target words 0-16 were /s/ target words and 17-32 were /ʃ/ target words. A line is also plotted at 0.5 to aid in the visualisation of which continuum point is closest to this value.

Therefore, closely linked to these individual target differences was the variation by phoneme. Overall, differences between targets with an expected /s/ and /ʃ/ phoneme could be seen when collapsed across targets. The pattern was a repetition of all previous experiments when broken down by target word: The /s/ containing target shows lower proportions of 'word' responses continuum medially compared to the /ʃ/ targets. The pattern for the current pre-test seen in Figure 7.2 was not as pronounced when collapsed across targets as in previous experiments, with a difference of only 0.1 between the respective average proportions of 0.68 for expected /s/ targets and 0.78 for expected /ʃ/ targets. In Experiment 5 the average proportion of 'word' responses for the /s/ target word, 'Impressive', for the three continuum points being tested here (2, 3 & 4) was 0.35, whilst the average across all /s/ targets for the current study was almost double at 0.68. The /ʃ/ target, 'Condition'

had an average of 0.64 across the same continuum points for Experiment 5.

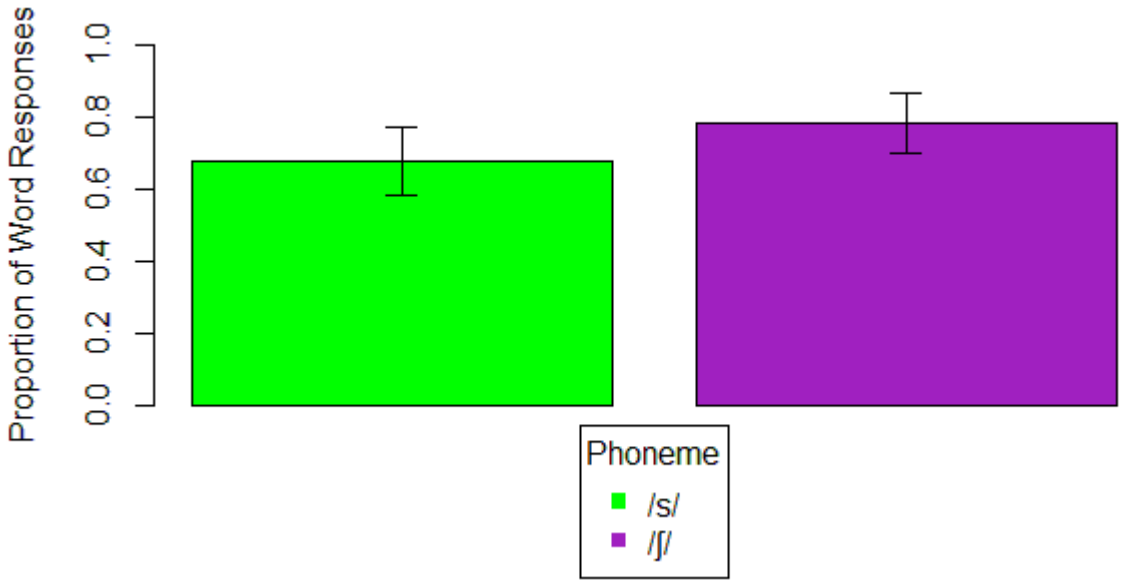


Figure 7.2- The Proportion of 'word' responses for Pre-Test 1 experimental items broken down by the 'Phoneme' (/s/ & /ʃ/) that would be expected in a target word.

The primary focus of the pre-test was to select the variant of target word that coupled with one of the three medial continuum points resulted in the proportion of 'words' value closest to 0.5. Table 7.1 shows the proportion of lexical decisions and the continuum point chosen for each target.

We created a set of criterion for selecting the continuum point for each target: If there was a 0.5 proportion present then this continuum point was chosen; If there was more than a single point at which the 0.5 value occurred then there was preference for the balanced fricative variation created at point 3; If there was not a value of 0.5 for proportion of 'word' responses then the continuum point with the closest value was chosen; if two of the continuum points had proportion values that were of equal magnitude away from 0.5 but in the same direction then we preferred the balanced fricative variation of point 3 and if two of the continuum points had proportion values that were of equal magnitude away from 0.5 but in different directions then

we preferred the value that was above 0.5. Across all target words the average 'word' response proportion value was 0.6. This was above our ideal value of 0.5 but as described above we preferred values above 0.5 and there was a tendency in the data for values above 0.5.

Due to having selected only a third of the possible continuum points tested, the average 'word' proportions between the phonemes for the continuum points chosen to be used in the main experiment showed values slightly about our target of 0.5: 0.59 for expected /s/ target words and 0.60 for expected /ʃ/ target words. There was no reliable difference between the targets expected to contain /s/ or /ʃ/.

The continuum point chosen was important because this describes the constituent phoneme structure of the fricative variation. The most chosen point was continuum 3 selected for 44% of target words; 57% of the points were for /s/ targets and 43% for /ʃ/ targets. Continuum Point 2 was chosen for 38% of targets: 83% of which were /ʃ/ targets and only 17% /s/ targets. The remaining 19% targets were selected with Continuum point 4: All were /s/ targets. To put these continuum points into context, continuum point 2 was made of majority /s/ phoneme and point 4 was majority /ʃ/.

7.4 Experiment 6: Speech Perception, Focus and Disfluency

In the current study we asked how focused listener attention and disfluency affected a lexical decision task when there was pronunciation variation at a phonemic level in a number of target words, with a view to investigate the attentional account of disfluency processing. With the increase in the number of target words tested with only medial continuum points and new materials we set out to see how this altered the results seen in the previous experiments.

Item	*Target*	*Phoneme*	*Continuum Point*	*Proportion of Word Responses*
1	messy	/s/	3	0.5
2	fossil	/s/	3	0.63
3	gossip	/s/	3	0.5
4	essay	/s/	4	0.5
5	lesson	/s/	2	0.63
6	message	/s/	4	0.75
7	classic	/s/	4	0.5
8	crossing	/s/	4	0.63
9	kissing	/s/	3	0.63
10	ascend	/s/	3	0.5
11	assume	/s/	3	0.75
12	dressing	/s/	3	0.5
13	racing	/s/	3	0.63
14	guessing	/s/	2	0.75
15	passing	/s/	4	0.5
16	receipt	/s/	4	0.5
17	crushing	/ʃ/	2	0.5
18	lotion	/ʃ/	2	0.63
19	fishing	/ʃ/	2	0.5
20	caution	/ʃ/	3	0.63
21	usher	/ʃ/	2	0.5
22	mission	/ʃ/	2	0.5
23	pressure	/ʃ/	2	0.75
24	machine	/ʃ/	2	0.5
25	nation	/ʃ/	3	0.63
26	motion	/ʃ/	3	0.75
27	wishing	/ʃ/	3	0.63
28	fashion	/ʃ/	2	0.75
29	cashier	/ʃ/	2	0.5
30	ocean	/ʃ/	3	0.88
31	cashew	/ʃ/	2	0.5
32	initial	/ʃ/	3	0.75

Table 7.1- Experimental Targets with Phoneme, selected Continuum Point and Proportion of Word Responses.

7.4.1 Disfluency Creation

Disfluency is a key part of the current study, so it was important to create tokens of disfluency that closely matched the phenomenon when occurring in natural speech and that could generalise to instances of disfluency employed in other studies.

Disfluency always occurred in the same location throughout the experiment, pre-final target word. The motivation for this location was that effects had been seen in the previous experiment with this disfluency position, proving the efficacy of this location. The current paradigm followed the same trial structure as the previous study, so disfluency was expected to act in a similar manner. In Experiment 5, we used two variants of a filled pause: 'uh' and 'um'. Due to a lack of difference between these variants in the previous study when analysed together, we only employed, 'um' in the current paradigm. The average duration of the disfluencies in the current study was 604ms (SD: 93ms). This was similar to the duration of the 'um' filled pauses (614ms) used in the previous study.

All of the sentence place holders had a matching disfluent version that was recorded separately. The average duration of the fluent sentence place holders was 1600ms and the average duration of the matched disfluent sentence place holders was 2491ms: A difference of 891ms. This difference was longer than the fluent presentation of the sentences with the addition of the average disfluency. This discrepancy was likely down to natural variance in production and additional pauses that surround the disfluency. This phenomenon was observed in Experiment 5. However, the difference was relatively small: 287ms or 18% of the average fluent place holder duration. This short increase in duration was unlikely to have affected participants' lexicality judgments, as the extra time did not give them any advantage in the task.

The disfluent sentence holders were recorded in the same session and manner as the fluent sentence place holders and all targets detailed above. The disfluent sentential

contexts were always produced with the token “pen” as the final referent, keeping the effects of co-articulation and prosody between the sentence and experimental target words constant throughout the experiment. The auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. All materials were produced by a native British English speaker. The speaker was instructed to produce the materials in a naturalistic manner and the materials were repeated until this was achieved.

7.4.2 Participants

A total of 24 students from the University of Edinburgh participated for a reward of £4.50 upon successful completion of the study. Participants self-reported that they were native speakers of English and had no speech or hearing difficulties. Participants who had taken part in either the pre-test for the current study, the pre-tests or main studies for previous speech perception Experiments 4 & 5 were excluded from taking part in the current study.

7.4.3 Design and Materials

Each trial was made up of a place holder sentence followed by a target word that participants then had to make a lexicality judgement on. There were two types of target word: experimental targets and filler targets. The experimental targets were described above in the target word section. In summary, there were 32 experimental targets: 16 /s/ target words and 16/ʃ/ target words. A complete list of the experimental targets and the continuum point used for that target can be found in Table 7.5. There were 64 individual two-syllable filler targets: 32 'word' and 32 'non-word' fillers. In each category of filler target ('word'/'non-word') there were 16 /s/ targets and 16 /ʃ/ targets. All fillers were used in the Pre-Test. The 'non-word' fillers had the naturally occurring fricative phoneme exchanged for the fricative phoneme at the opposite end of the continuum for normal usage: /s/ and /ʃ/ were swapped.

There were 32 place holder sentences included to increase the ecological validity of the task by lowering the repetition of hearing a place holder in comparison to the

previous study. There was a fluent and disfluent version of each of the 16 sentence place holders. They were short and context neutral, so that participants were not anticipating any certain entity and so that they could accommodate all experimental targets and filler targets. An example was, “She remembered to say…”. The experimental instructions gave the place holders context by stating that they related to a native British English speaker giving instructions about a word list. All sentences were between 5-12 syllables long and finished with either “say” or “be”. They followed a similar structure so that the effect of the place holder sentence would be minimised.

A complete experiment consisted of 96 trials that were made up of 32 trials with experimental targets and 64 trials with filler targets. Each experimental target only occurred with one of the fluency versions of the sentence place holders. Half of the experimental targets were paired with disfluent presentations. Both /s/ and /ʃ/ phoneme target words were presented with 8 fluent and 8 disfluent sentence place holders. Each filler target was only presented with one sentence place holder and 16 of each category of filler targets ('word'/'non-word') were presented with fluent place holders and 16 with disfluent sentence contexts. Again, there was an equal number (8) of fluent and disfluent presentations for filler targets containing each phoneme (/s/ & /ʃ/). Participants heard each sentence place holder a total of 6 times: 3 fluent and 3 disfluent presentations.

The experimental targets made up 33% of trials with fillers presented for the remaining 66% of trials. This was higher than the percentages of experimental targets (14%) and fillers (86%) seen in Experiment 2 in Pitt and Szostak (2012) but there was no repetition of targets. This meant the experimental targets should have been effectively disguised to stop participants being able to differentiate them from the filler targets. The experiment was presented as a single block with trials randomly presented within the block.

The instructions carried the focused attention manipulation, either Focused or Unfocused. Participants only ever saw one set. The instructions accounted for the

pronunciation variation as 'mistakes'. The instructions used for the current study were matched to those used in Experiments 4 & 5. The 'Focused' condition instructions alerted participants that possible mistake would always be in the final word and that changes could be small and would be sound based and took place at the start of the final word. We added additional focus to the /s/ and /ʃ/ phonemes by defining the task in greater details using these phonemes: " You may for example hear the speaker saying '*sh*' in the middle of a word when they mean '*s*', in which case they may have mistakenly produced a sound that isn't a word". However, in the 'Unfocused' condition the instructions simply stated that there could be some mistakes and did not emphasise which phoneme sound or the phoneme location within the word where mistakes could occur. The instructions here diverged from Pitt and Szostak (2012) because of the differing task demands of having a place holder sentence, which meant that we had to identify the final word as being the token that participants had to make a lexical decision on. Both sets of instructions can be seen in full in Appendix A. Half of the participants saw the 'focused' instructions with the remaining participants seeing the 'unfocused' instructions.

Comprehension questions were included after 21% of trials, so that engagement with the task could be gauged throughout the task. These questions only followed trials which contained a 'word' filler target but the place holder preceding the target could have been either fluent or disfluent. Equal numbers of comprehension questions (10) followed fluent and disfluent presentations. The comprehension questions asked participants to select between two choices: the target they had just heard or a competitor word. The competitors were phonetically or semantically similar to the target word heard, for example, a filler target was "Notion" and the competitor target for this trial was "Noting".

All auditory stimuli were recorded at a University of Edinburgh studio facility by an engineer with the author present. The filler targets were recorded using the same procedure described above. The recording process for the sentence place holders is also detailed above in the disfluency creation section.

7.4.4 Apparatus and Procedure

The apparatus and procedure followed that described in Experiment 5 but due to paradigmatic variations there were some minor differences: Due to there being only one block participants did not have any breaks and the duration of the current experiment was much shorter at approximately 30 minutes.

7.4.5 Measures

The measures that were used were the proportion of word responses for each continuum point and the percentage of comprehension question that were answered correctly.

7.4.6 Analyses

We analysed participant's lexicality judgements. Our primary focus was the proportion of 'word' responses and how this data looked when broken down by Focus, Disfluency and Phoneme conditions. All analyses were only undertaken on the experimental target data, filler targets were removed. We excluded trials where participants did not make any selection, leading the trial to time out. This accounted for 4% of all trials. For each trial, if the participant selected a word we coded this as 1 and if a non-word then this was coded as a 0. Due to our dependent variable being binomial (whether a participant judged a target as a word or not), we employed the same analyses as in the previous studies (Experiments 2-5): a linear mixed-effects regression model with empirical logit transformed proportion data. This model was 'maximally specified' with both random intercepts and slopes, as well as their correlations varying by participants, as suggested by Barr, Levy, Scheepers, and Tily (2013). The reasoning for the choice of an empirical logit transformation was that we expected that at the Continuum endpoints there would be either a lot of 0s but few 1s, or vice versa. When this occurs logistic regressions tend to have problems converging. This problem is minimised when an empirical logit transformation is employed. The predictors we used in the analyses were Focus (Focused and Unfocused) which was between participants and Disfluency and Phoneme (/s/ & /ʃ/)

which were within participants. The length of experiment did not warrant splitting into two lists and the duration of experiment was short enough to expect that participants could concentrate for the complete experiment without experiencing difficulty. So for the current study there were no Half or List predictors as had been employed in previous experiments.

The comprehension question data was used as a check throughout the experiment: If a participant was consistently answering comprehension questions wrong then we would question the validity of their data. For each comprehension question we coded 1 for a correct answer and 0 for an incorrect answer and then we created a percentage for each participant, based on the number of correct responses.

7.5 Results

We first present the comprehension question results, as this could have affected the data taken forward into the analysis of the lexicality judgements for the main experiment. The results of the lexicality judgement analyses are presented following this. All lexicality judgments were analysed in R (R Development Core Team, 2014) using the lme4 package (Version 0.999999-0, Bates, Maechler & Bolker, 2014), p values were calculated using the lmerTest package (Version 1.2-0, Kuznetsova, Brockhoff & Bojesen, 2013).

7.5.1 Comprehension Questions

As described above, we wanted to check participants' answers to the comprehension questions to decide whether the rest of their data should be included in the analyses. One participant's comprehension data was excluded from the analyses below, as they did not answer any of the comprehension questions. They misunderstood the comprehension question task and informed the experimenter of their confusion at the end of the experiment. However, this participant responded to the rest of the filler trials in the expected way. On this basis their data was included in the main

analyses, as there was a systematic reason for their 0% response rate of the comprehension questions rather than the lack of attention that we hoped to guard against.

For the remaining comprehension question data, we excluded trials where the comprehension question trials had timed out and not been answered. This accounted for 12% of all trials. This is higher than in previous experiments but there were fewer trials compared to previous studies, so fewer timed out questions would be needed for a higher percentage. With this data removed, the lowest comprehension question score was 93% of comprehension questions answered correctly, showing that participants were answering the questions consistently correct. On this basis all participants' data was included in the main analyses.

7.5.2 Proportion of Lexical Responses

The current study was investigating whether there was for an effect of 'focused' attention and disfluency on the proportion of 'word' responses made by participants. Figure 7.3 shows the proportion of lexical decisions collapsed across Disfluency and Phoneme conditions but broken down by Focus condition, into 'Focused' and 'Unfocused' instruction conditions. The pattern observed was clear: The 'Focused' condition (0.54) showed a decrease of 0.12 in the proportion of 'word' responses compared to the 'Unfocused' condition (0.66). The Focus conditions showed a robust difference in the linear mixed model described above (β= -0.30, SE=0.13, t=2.38, p=0.027). This was a repetition of the 'Focused' attention effect seen in Experiment 4 and a replication of the effect seen at the same continuum point in Pitt and Szostak (2012): 'Focused' Attention increases participants' sensitivity to the fricative variation, resulting in reduced lexical judgements.

Although the current study was comparable to the Pitt and Szostak (2012) paradigm, albeit with a preceding sentence placeholder, and our previous experiments, comparing the proportion values for the current study against these is complex, as

the values seen here are collapsed across different phoneme variation and there was no clear continuum point to compare against from the other studies. Therefore, we do not compare the proportions of 'word' responses for the predictors here with previous variants.

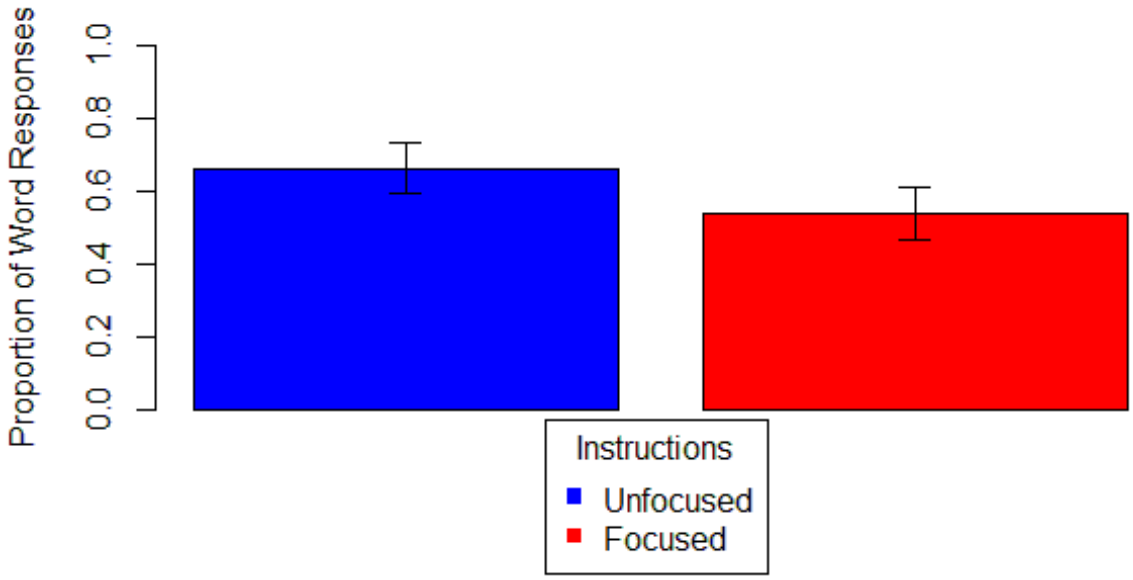


Figure 7.3- The proportion of 'word' responses by Instruction type (Focused/Unfocused).

Investigating the impact of disfluency was the central focus of the current study. Figure 7.4 shows the lexical response data broken down by Fluency condition but collapsed across Focus and Phoneme conditions.

There was a slight increase of 0.04 proportions of 'word' responses for the Disfluent presentations (0.62) over the Fluent variants (0.58). This difference was smaller than seen in Experiment 5 at Point 3 but consistent with the direction of the effects seen. However, this small difference was robust with a main effect observed for the Fluency predictor in our linear mixed effects model (β= 0.2, SE=0.09, t=2.31, p=0.031). This was a repetition of the disfluency effect seen in Experiment 5 at Point 3. The incidence of Disfluency immediately prior to a target word increased participants'

'word' responses. In short, Disfluency made participants more accommodating of word medial fricative variation.

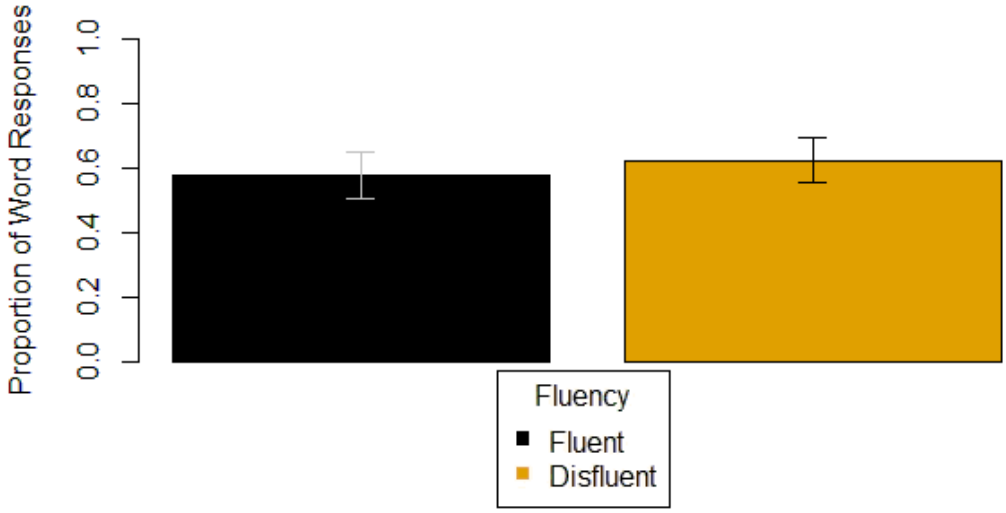


Figure 7.4- The proportion of 'word' responses by Fluency Condition (Fluent/Disfluent).

In all previous speech perception experiments reported in the current thesis there has been a large effect for the Phoneme predictor (reported as Target in previous experiments due to only having a single word pair), with the target word with an expected /ʃ/ phoneme always exhibiting increased proportions of 'word' responses. This pattern was most pronounced Continuum medially. This pattern was repeated for the current study; targets expected to contain the /s/ phoneme (0.39) showed a reduction of 0.42 in proportion of 'word' responses in comparison to targets expected to contain the /ʃ/ phoneme (0.81) as seen in Figure 7.5.

This was a large difference and over 10 times bigger than the difference seen previously for the Disfluency predictor. This unsurprisingly led to a reliable main effect in our mixed effects model for the Phoneme predictor (β= 1.05, SE=0.16, t=6.34, $p < 0.001$). This main effect repeated those seen by phoneme with and without the influence of disfluency in previous studies. Participants were more forgiving of fricative variation in target words when they are expecting to hear an /ʃ/ phoneme than when they are expecting to hear an /s/ phoneme.

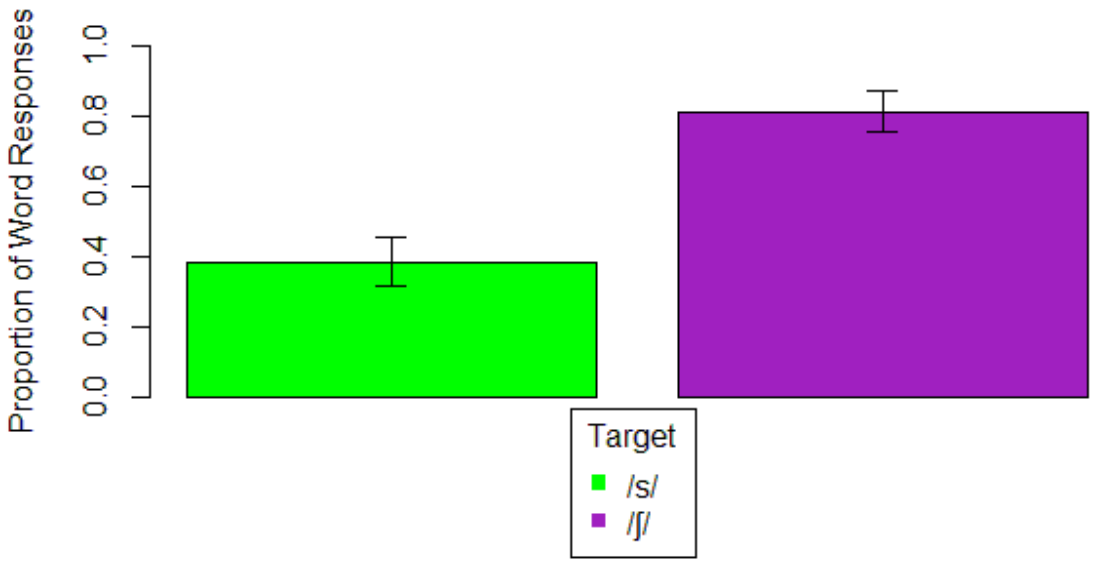


Figure 7.5- The proportion of 'word' responses by expected Phoneme in the Target words (/s/ or /ʃ/).

Next we focused on breaking down the responses by more than one predictor. Figure 7.6 shows the proportion of lexicality judgments when the data is broken down by both of our primary interest predictors, Fluency condition (Fluent/Disfluent) and Focus condition ('Focused'/'Unfocused').

Both 'Focused' conditions (Disfluent presentation: 0.58 / Fluent presentation: 0.51) showed a reduction in 'word' responses compared to their respective 'Unfocused' variants (Disfluent presentation: 0.68 / Fluent presentation: 0.65). The fluent presentations had an increased difference of 0.14 between their focus conditions and the incidence of Disfluent sentence place holders reduced this gap to 0.1. Both of the disfluent conditions show slightly increased proportion of 'word' responses compared to their fluent equivalent. In the Disfluent and Focused Condition (0.57) it had an increase of 0.06 in proportion of 'word' responses compared to the Fluent and Focused condition (0.51); This gap was halved to 0.03 between the Disfluent and Unfocused (0.68) over the Fluent and Unfocused pairing (0.65).

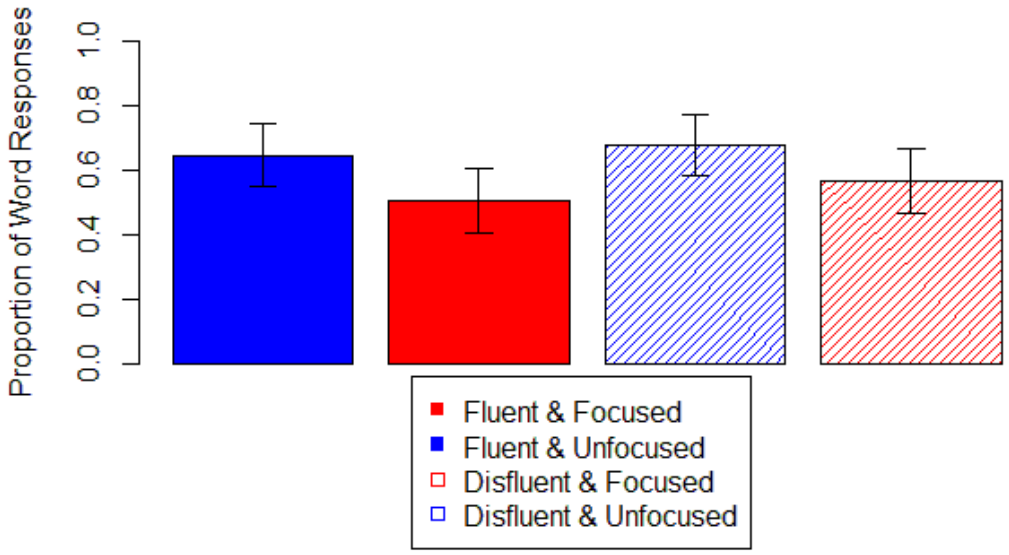


Figure 7.6- The proportion of 'word' responses by Focus Condition ('Focused'/'Unfocused') and by Fluency Condition (Fluent/Disfluent).

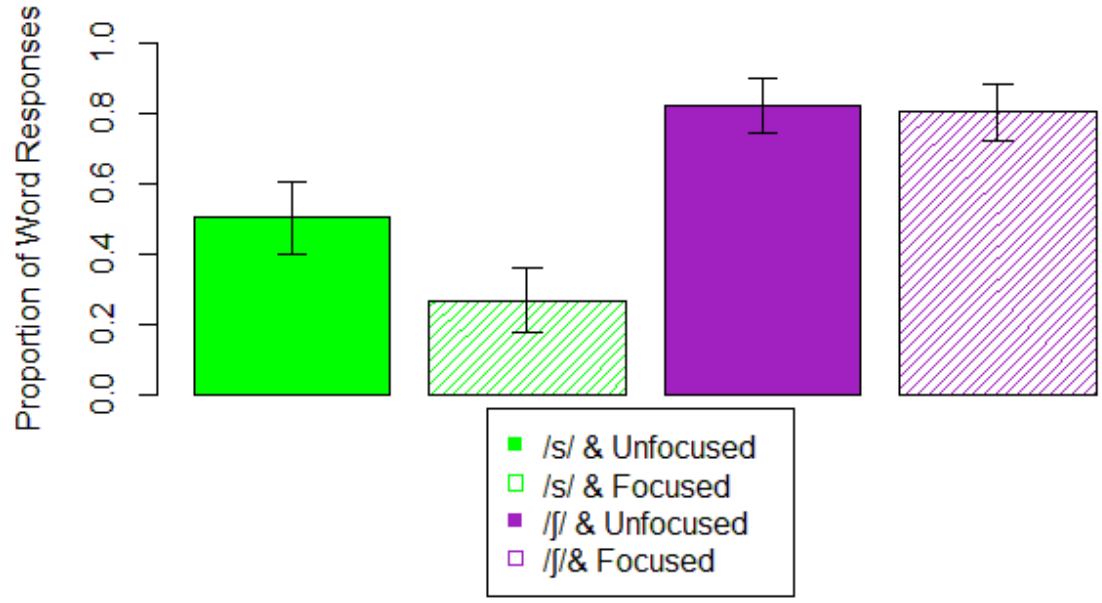


Figure 7.7- The proportion of 'word' responses by expected Phoneme in the Target words (/s/ or /ʃ/) and by Focus Condition ('Focused'/'Unfocused').

The patterns observed here reflect both of the main effects seen for Fluency and Focus conditions above: Focused attention reduced lexicality judgements regardless of the fluency of the place holder, although a bigger reduction was seen for Fluent presentations; Disfluent presentations of sentence place holders led to increased

proportion of ‘word’ responses compared to Fluent variants over both Focus conditions, although a larger gap was seen for the 'Focused' condition. With the effects of both Fluency and Focus predictors consistent, there was unlikely to be an interaction effect and none was observed (t <1).

As detailed above, large effects were observed for the Phoneme predictor and we wanted to investigate how this effect broke down when further divided by both Focus condition and Fluency condition. Firstly, we looked at 'Focused' attention, Figure 7.7 shows the proportion of lexicality judgements by both the Phoneme and Focus predictors. A repetition of the main effect of Phoneme predictor observed above is clear: Both Focus conditions for the targets expected to contain /s/ are lower than both conditions for the /ʃ/ targets. Comparing the Focus conditions by Phoneme revealed a large reduction of 0.54 between the 'Focused' conditions (/s/: 0.27 and /ʃ/: 0.81) and a reduced gap of 0.32 between the 'Unfocused' conditions (/s/: 0.50 and /ʃ/: 0.82). For both phonemes, 'Focused' attention produced a reduction in the proportion of ‘word’ responses compared to the matched 'Unfocused' condition. There was a much larger variation between Focus conditions for targets expected to contain /s/ (0.23) compared to those expected to contain /ʃ/ (0.01). This was further evidence of the 'Focused' attention effect seen in Pitt and Szostak (2012) and Experiment 4. However, this attentional effect was considerably reduced in the targets expected to contain /ʃ/. There was no interaction effect between the Phoneme and Focus predictors (t <1.4).

Next we explored the response data when broken down by both Phoneme and Disfluency predictors. Figure 9.11 shows the lexical decision proportions for both Fluency conditions and Phonemes. The expected main effect of Phoneme was clearly visible with /s/ phoneme targets lower than either of the expected /ʃ/ targets. The Fluent presentations in the /s/ targets (0.39) showed a reduction of 0.37 compared to the matched Fluency condition in the /ʃ/ targets (0.76) and the disfluent variant for the /s/ target (0.38) showed an increased reduction of 0.49 compared to the disfluent presentation for the /ʃ/ target (0.87).

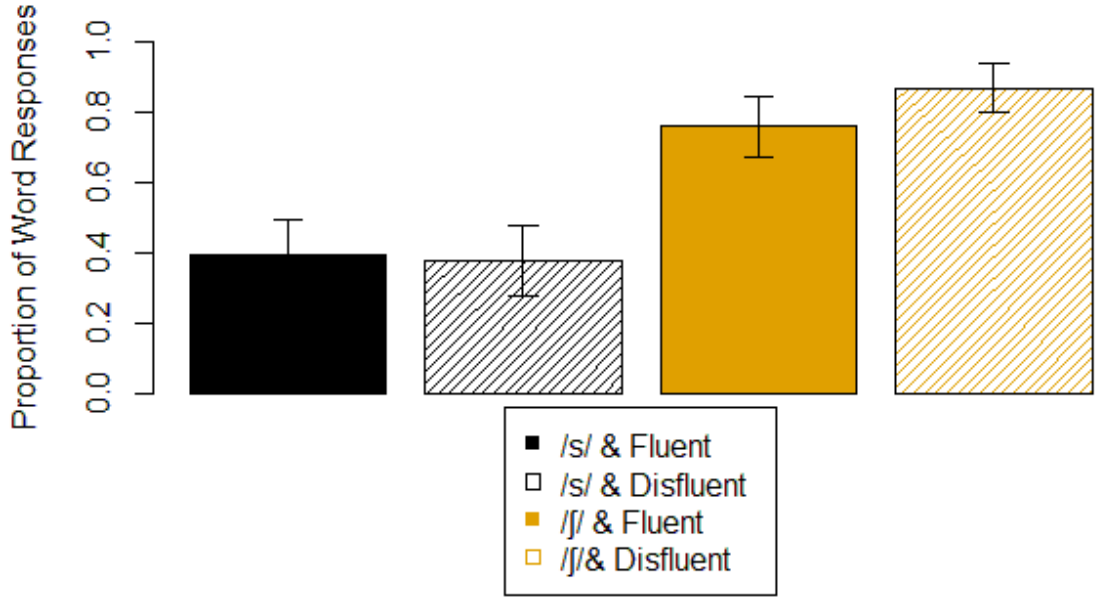


Figure 7.8- The proportion of 'word' responses by expected Phoneme in the Target words (/s/ or /ʃ/) and by Fluency Condition (Fluent/Disfluent).

The main effect of Disfluency detailed above was not clear when broken down by phoneme. There was a repetition in the targets expected to contain the /ʃ/ phoneme with the disfluent presentation (0.87) showing an increase of 0.11 proportions of 'word' responses over the fluent variant (0.76); however for the targets expected to contain an /s/ phoneme the incidence of disfluency (0.38) actually led to a small decrease of 0.01 below the fluent presentations (0.39). The disfluency effect was only seen for one of the Phoneme predictors and this was reflected in an interaction effect between Disfluency and Phoneme (β= 0.18, SE= 0.08, t=2.37, p=0.027).

The interaction between Phoneme, Disfluency and Focus was not graphed here due to no effects being observed for this combination of predictors (t <1).

7.6 Discussion

In the current study we set out to investigate the attentional account of disfluency by creating a new paradigm that built on the results of the previous speech perception studies by focusing continuum medially but with an increased number of target words, so that the results could be generalised, rather than being susceptible to the individual variance that may have affected the pairs of target words used in previous paradigms. We tested the impact of focused attention and the inclusion of disfluency housed within a sentence place holder on participants' lexicality judgements for a target that contained phoneme variation.

The focused attention manipulation used instructional conditions, as in Pitt and Szostak (2012) and our previous studies. The results observed here supported our prediction with the focused instructions causing a reliable reduction over the unfocused instruction condition. This provides evidence that focused attention impacted the current paradigm in the same manner as in Experiment 4 at continuum point 3 and Pitt and Szostak (2012). The results are not directly comparable, as in the current study we employed varying phoneme variation in a word medial location that does not directly align with any point in either Experiment 4 or the Pitt and Szostak paper. However, this creates a more compelling effect as it is spread across a number of continuum medial points employed in these studies. Additionally, in the Pitt and Szostak paper, variance was seen across the continuum, so the current focused attention results support this. In Experiment 5 there was no attentional effect seen but the current results that tested the impact of focused attention across an increased number of target words are more likely to generalise to reflect reality, as they are less susceptible to individual variance than the pair of target words used in the previous studies. The attention effect did not appear to be balanced between /s/ and /ʃ/ target words: the /s/ target words showed a much larger decrease in the proportion of 'word' values when broken down by Focus condition, whereas, the /ʃ/ target words were equivalent with a decrease of only 0.01 proportions of 'word'

responses. However, no interaction effect between Phoneme and Focus condition was observed and a main effect remained.

The direction of the impact of disfluency in Experiment 5 was unexpected, the previous results showed that the inclusion of disfluency made participants more accommodating of word medial fricative variation at Continuum Point 3. This effect was replicated in the current study, with a main effect of disfluency observed such that the disfluent condition caused a small but reliable increase in the proportion of 'word' responses over the fluent condition. However, when the responses were additionally broken down by expected phoneme the disfluency effect did not create equal differences between fluency condition in both /s/ and /ʃ/ expected target words. The fluency effect was driven by the larger differences seen for /ʃ/ target words, whereas, for the /s/ target words the number of 'word' responses were similar across fluency conditions. This resulted in an interaction effect being seen.

Previously in Experiment 5, when the disfluent and focused attention conditions coincided it led to a reduction in the magnitude of variance away from the Fluent conditions, compared to the responses broken down by both disfluent and unfocused condition. If the results were to match this we would have expected the Focused attention conditions to show a reduced disfluency effect. A different pattern was observed in the current study, with both of the single predictor effects being preserved: Focused attention reduced lexicality judgements regardless of the fluency of the place holder; Disfluent presentations of sentence place holders led to increased proportion of 'word' responses compared to Fluent variants over both Focus conditions.

In Experiment 5 we highlighted that we created explicit differences in task demands between the Focus conditions but disfluency may have generated a different set of task demands that need not necessarily facilitate increased awareness of fine acoustic detail. Here the results suggested that both the disfluency and focused attention effect can work concurrently.

There were an increased amount of target words used in the current study and these were updated from the previous study. In our pre-test the average 'word' response values between target words containing each phoneme were relatively equivalent, with a gap of only 0.01. Yet there was still large variance between those expected to contain /s/ and those expected to contain /ʃ/. This provides more evidence that participants seem to be more accommodating of fricative variation when they expect to hear an /ʃ/ phoneme compared to when they expect to hear an /s/. It follows that the consistent differences between targets is harder to account for aside from factors that we cannot investigate from the current results such as the difference in the acoustic characteristics of the stimuli from Pitt and Szostak's study or differences between British and American English or sensitivity to variation that may vary between groups of participants from these respective countries. If our continuum was more natural sounding than Pitt and Szostak's then this could have influenced participants' sensitivity to the phoneme variation heard. It is worth noting that we do not know how Pitt and Szostak's pattern of results breaks down by expected phoneme but regardless of any differences between targets, their results generalised across the full continuum.

It was interesting that focused attention and disfluency effects appeared to be having differential effects on target words, with focused attention showing larger differences for the /s/ target words and disfluency showing larger differences for the /ʃ/ target words. A possible explanation for these effects relates to participants being more accommodating of fricative variance when they are expecting to hear an /ʃ/ phoneme, as opposed to an /s/ phoneme due to the experience of these phonemes that the listeners gained in the preceding sentence place holders. The /s/ phoneme occurs more frequently than /ʃ/ in English in general and that pattern was repeated in the current study. Across the sentence place holders there was 14 instances of /s/ phonemes heard, compared to only 4 instances of /ʃ/. This created unbalanced experience of the speaker producing these phonemes in other contexts that could have influenced the listeners' sensitivity to what was an acceptable /s/ versus what was an acceptable /ʃ/. This explanation would account for why there was limited

differences seen when the targets were presented as single words in the pre-test, compared to the reliable differences observed between targets in the main study when they were heard following a sentence place holder. In all of the speech perception experiments contained in the current thesis, at continuum medial points /ʃ/ target words have had increased proportions of 'word' responses compared to the /s/ targets.

It follows that we could categorise the /ʃ/ targets as consistently being heard as more 'word' like, this could have been a reflection of increased lexical bias for the same fricative variation for these targets, compared to the /s/ targets. Therefore, if disfluency was having a facilitative effect on lexical bias, it would act more on the /ʃ/ targets, compared to the /s/ targets for the same continuum point, which would explain the increased effect of fluency condition for the /ʃ/ targets. This pattern of results was reversed for focus conditions, with focused attention exerting a much larger effect on the /s/ targets, compared to the /ʃ/ targets. The focused attention was acting in the predicted direction for the /s/ targets, resulting in a decrease compared to the unfocused instructions but there was only a very small reduction seen between for the /ʃ/ targets but this could be for the same reason.

A possible explanation again stems from participant having variable lexical bias for each phoneme sound, perhaps as a result of varying exposure to each phoneme sound in the sentence place holders. For the /s/ phonemes there was a reduced lexical bias which allows the focused attention condition to receive more information from the bottom-up acoustic information resulting in the reduction in proportion of word values seen. However, for the /ʃ/ targets there was a stronger lexical bias which means that the focused attention may still have been driving participants to employ more bottom-up acoustic information but it has to overcome greater top-down lexical activation, so the effect for these targets is reduced, leading to the smaller effects observed. This explanation is supported by the results observed in Pitt and Szostal (2012) who show reduced variance towards the 'word' end of the continuum.

As noted in Experiment 5, there may be a cognitive load explanation for the varying impact of lexical bias on participants' 'word' responses. If after encountering disfluency there was an increase in attention, this could lead to an increase in cognitive load experienced by the participant. Under cognitive load participants have been shown to rely more on lexical influences (Mattys & Wiget, 2011). It follows that for /ʃ/ targets, the strength of lexical bias before encountering disfluency or focused attention was greater than for the /s/ targets, which could explain the variable impact seen between disfluency and focused attention on these different targets, as these predictors were working on different base lexical biases for the /ʃ/ targets compared to /s/ equivalents. This explanation again points to the variable role of task, attention and disfluency in impacting sensitivity to lexical bias, which in turn points to the balance of bottom-up and top-down processing being crucial in both speech perception and disfluency processing. This is explored more below in the general discussion.

In terms of the attentional account of disfluency that the current study set out to investigate, the paradigm here was the most reliable as it was the best replication of the attentional effect seen in the Pitt and Szostak study (2012) and a disfluency effect that generalises across multiple target words, as opposed to just a single pair. The effect observed showed that encountering a disfluency made participants more accommodating of the phoneme variation compared to the fluent condition. This provides further evidence against our original hypothesis that if following disfluency there was heightened attention then this would be reflected in a reduction in the proportion of 'word' responses; similar to that seen for an instructional manipulation of focused attention that was also observed in the current study.

The disfluency effect seen here is not necessarily incompatible with the attentional modulation predicted as we have noted in Experiment 5 that there may have been varying task demands for the fluency conditions when compared to the focus conditions. Meaning that the disfluency could still be driving heightened attention but this may not to be to the incoming signal, as in the focused attention condition,

but instead increasing top-down lexical bias due to there being no necessary interest in fine acoustic detail for disfluent productions, aside from when co-occurring with the focused attention instruction condition. The pattern of results seen for the current study support this proposal, as when focus and fluency conditions overlap, both main effects continue to be observed. This is discussed further in the general discussion.

Taken together, these results were more reliable than those seen previously as they generalised across multiple target words. Both an effect of disfluency and focused attention were seen, acting in opposite directions. Disfluency made participants more accommodating of the fricative variation, whereas, focused attention made them more sensitive to the phoneme variation. There were still clear differences in target by whether a participants expected to hear /s/ or /ʃ/, with the disfluency and attentional effects impacting each set of targets differently. We propose that this stems from different lexical bias created between target words that contain /s/ opposed to those that contain /ʃ/, possibly as a consequence of a listeners' experience with each phoneme in the proceeding sentence place holders. Speaking to our central focus of investigating the attentional account of disfluency, the current study replicated the effect seen continuum medially in the previous experiment. We suggest that disfluency was still generating heightened attention but this was acting to increase top-down processing and enhancing the lexical bias experienced by target words, as a consequence of either or both of variable task demands or cognitive load.

CHAPTER 8

General Discussion

8.1 Chapter Overview

The experimental work in this thesis was intended to investigate the roles of prediction and attention in disfluency processing to better understand the underlying mechanisms that drive disfluency effects. We first tested both the predictional and attentional accounts of disfluency processing in an eye-tracking study. Following this, we focused on exploring the attentional account of disfluency using a speech perception paradigm, with a view to testing how disfluency might modulate listener attention. This chapter provides a summary of the findings across these studies and discusses the implications for disfluency processing. Finally, we consider what can be concluded from the current findings and how the methods of working employed in the current thesis may inform future studies.

8.2 Interpretation of the findings

In the following sections, we first recap the discussion of Experiment 1, then we look at the role of attention across Experiments 2-6, before considering the implications these findings have for disfluency processing.

8.2.1 Which mechanisms drive disfluency processing?

A central focus of the thesis was to explore the underlying roles of prediction and attention in disfluency processing. There have been attempts to categorise and understand the underlying mechanisms that are responsible for a range of disfluency effects seen during comprehension. Different models of disfluency processing have been proposed to explain these effects that centre on the role of prediction and attention. Another focus of the thesis was trying to differentiate between these two accounts; the predictional and the attentional.

Disfluency has been observed to modulate predictive processing: Upon encountering a filled pause listeners show a bias towards unknown or discourse new referents (Arnold et al., 2007, 2004; Bosker et al., 2014; Heller et al., 2014). The predictional account of disfluency processing (e.g., Arnold et al., 2007; 2004; Heller et al., 2014) suggests that upon encountering disfluency, a listener infers the speaker to be experiencing difficulty. This difficulty can be driven by the situation of the speaker, with speakers tending to be more disfluent when they are experiencing cognitive load (Bortfeld et al., 2001; Brennan & Schober, 2001), increased difficulty in lexical retrieval, for example when trying to produce a word that is contextually unpredictable or of low frequency (Beattie & Butterworth, 1979). Listeners use this knowledge to build up patterns of disfluency distribution information that inform their expectations of upcoming content for a speaker.

The attentional account states that encountering disfluency causes listeners to abandon predictional processes, instead listeners employ heightened attentional resources, relying on bottom-up information, the incoming speech signal, to resolve the comprehension difficulty posed by the interruption to the speech, whilst the increased attentional resources allow quicker recognition of following linguistic content. This account has support from a number of disfluency based empirical findings (e.g., Collard et al., 2008; Fox Tree, 2001). These accounts are explored in detail in Chapter 2.

8.2.2 The role of prediction?

The findings of the current thesis on the role of prediction in disfluency processing were not consistent with the predictional viewpoint outlined above. Our first study, Experiment 1, used a visual world eye-tracking paradigm intended to differentiate between the predictional and attentional accounts. The two accounts predicted differing patterns of fixation behaviour towards 'predicted', 'competitor' and 'related' pictures in a visual scene following listeners encountering a disfluency. Under the predictional accounts, the listener would anticipate the upcoming object to be harder to access for the speaker and an increased proportion of looks to the

'competitor' picture would be expected, as it is the only other plausible object in the scene. However, the attentional account proposed that as the expectations of the upcoming content cease following a disfluency, then this will cause them to abandon or attenuate predictions that the sentence will end with the 'predicted' item. Instead, we would expect the sentence context to exert a weaker effect and this would increase fixations on *both* the 'competitor' item and the 'related' item. The main experiment provided some unexpected results, as the fixation behaviour observed was not predicted under either account. Instead, following disfluency participants made an increased proportion of looks towards the 'predicted' item.

A potential cause of these findings is that participants may not have been sensitive to the disfluency employed in the main eye-tracking study. This was further investigated in a number of post-hoc tests but the results observed were again inconclusive. Taken together, these findings suggested that disfluency processing is flexible and dependant on the task being undertaken. We proposed a combined accounts that took elements of the both predictional and attentional accounts. The attentional account proposes that following a disfluency predictional processes are stopped and the additional attentional resources are used to focus on bottom-up processing to facilitate comprehension. However, clearly following disfluency participants are employing predictional processing based on the contextual fit of the sentence context. The Combined account proposes that the heightened attention seen following a disfluency (e.g, Collard, Corley, MacGregor, & Donaldson, 2008) can be complementary to other processing such as predictional effects (e.g., Arnold et al., 2007; Heller et al., 2014). The findings of Experiment 1 are discussed in-depth above in the Experiment 1's general discussion, 3.13.

8.2.3 The role of attention?

The focus of the remaining experiments (2-6) was to further test the role of attention in disfluency processing with a view to exploring the attentional account, outlined above. We used an extended version of the speech perception paradigm employed in Pitt & Szostak (2012) to test whether encountering disfluency was modulating an

increase in listener attention. The task participants had to undertake was a lexical decision task, they had to respond to target words that contained phoneme variation running along a 5 point /s/-/ʃ/ fricative continuum in a number of word locations: Word initial, medial and final. The targets that contained the fricative continuum created a continuum of stimuli that ranged from 'word' to 'Non-word'. The explicit attentional manipulation that Pitt and Szostak employed was the instruction condition that participants saw; "Participants given the focused instructions were informed that the ''s'' or ''sh'' letter sound in a particular word position could be ambiguous, and that they should listen closely so as to make the correct response" (Pitt & Szostak, 2012: 1229). The unfocused instructions did not give the target phoneme or location within a target word or not. For the studies in the current thesis (Experiments 2-6), we added a neutral sentence place holder to the Pitt and Szsostak single word design so that disfluency could easily be added to the study in a pre-target location. The Pitt and Szostak paradigm was chosen as it had already shown participants to be sensitive an effect of Focused attention. Their findings showed that attention was increasing participants' sensitivity to the phoneme variation in the targets words, resulting in a decreased proportion of 'word' responses. We reasoned that if disfluency was causing heightened attention then in this paradigm the realisation of the results would be in a similar manner to the pattern of results already seen for the focused attention manipulation in the Pitt and Szostak study.

In Experiments 2 and 3 we tested our continuum of fricative variation in a word initial location. In these studies we did not include any disfluency in the paradigm, as we wanted to first establish that we could generate an effect of focused attention as seen in Pitt and Szostak (2012). However, we failed to find any reliable effect for by instruction condition, although we did identify some paradigmatic weaknesses that left the results from these studies subject to question: In Experiment 2, we had two target words and the /s/ variant, 'Sandcastle', contained a second /s/ sound. We argued that this was likely to have influenced participant's judgement of the lexicality of the target, especially when they were instructed to listen for that sound. In Experiment 3 the /s/ variant again showed a weakness of being a compound noun,

'Sandpit'. We suggested that this may have influenced the ease of lexicality decision for participants, as they knew they had to judge the final word and it may have not been clear whether they were judging 'Sandpit' or just 'pit'.

In Experiment 4, we updated the paradigm to account for the weaknesses seen in the previous experiments. We moved the fricative variation to a word medial location, as this was where the biggest effects had been observed by Pitt and Szostak (2012). This required us to select new target words to carry the new continuum of fricative variation. The findings from this study did conform to those of Pitt and Szostak with regard to the nature of the attentional effect across the continuum. However, we did find an effect of focused attention continuum medially, with consistent by-condition (Focused/Unfocused) differences observed at this midpoint. The attentional effect here acted in the direction predicted, with focused attention leading to a reduction in the proportion of 'word' responses, compared to the unfocused instruction condition.

This provided the result needed for the inclusion of disfluency and, therefore, in Experiment 5 half of the experimental trials were preceded with a disfluent sentence place holder. The results from this study did not support either a fluency or attentional effect across the whole continuum. The largest differences between these predictors were seen continuum medially and a reliable disfluency effect was seen for two consecutive continuum points. Interestingly, the disfluency effect here emerged at a different direction in each point, with the continuum point closer to the 'non-word' end of the continuum showing a reduction in 'word' responses following a disfluency, whereas, at the midpoint disfluency was making participants more accommodating of the fricative variation. There was no focused attention effect seen continuum medially for this study, which was not consistent with the observation of a reliable effect at this point in Experiment 4. The presence of disfluency in Experiment 5 appeared to have an experiment wide effect that counteracted the impact of focused attention seen previously in Experiment 4. These results for Experiment 5 suggested that disfluency was having a variable impact based on the

phoneme variation being heard. We argued that the findings supported the impact of disfluency and the role of attention as being modulated by lexical bias, task demands and the cognitive load experienced by the user. This is discussed further below. We suggest here that due to the effects only being seen across two target words, they are susceptible to variance that may stem from the individual words tested. We decided that for a thorough investigation of the effect of focused attention and disfluency within this paradigm that we needed these effects to generalise across target words.

Our final speech perception study, Experiment 6, made use of 32 targets that each contained only a single word medial continuum point. The phoneme variation chosen was the point closest to a value of 0.5 proportion of word responses from the three continuum medial points used in Experiments 4 and 5 during a pre-test. These changes to the paradigm centred on the area of the continuum where the largest variance had consistently been seen in the previous experiments, whilst also creating an ambiguous phoneme. Additionally we selected the target closest to this 0.5 value as this meant lexical bias should have been balanced, as each point was theoretically equal in ratings of word responses. The pre-test supported this with minimal variance between the lexicality ratings for /s/ and /ʃ/ targets.

The results of this study showed reliable main effects for both the Focus and Fluency conditions. The effect of Focus was consistent with that observed in Experiment 4, with a reduction seen for participants' 'word' responses in the Focused attention condition. The incidence of disfluency in the preceding sentence place holder made participants more accommodating of the fricative variation, resulting in an increased proportion of 'word' responses, compared to the values seen for the fluent condition. This pattern of results for the fluency conditions matched the effect seen at the Continuum midpoint in Experiment 5. These effects remained when the responses were broken down by both Focus and Fluency. This suggests that both the disfluency and focused attention effects are additive and can work concurrently. The results seen here do not support the predicted disfluency effect proposed by the

attentional account of disfluency processing. However, the results observed here can be reconciled with the Combined Account proposed above. This is discussed further below.

Across all of the speech perception studies in the thesis, there was a consistent effect of the phoneme that participants expected to hear in the target word: Participants showed a tendency for increased lexical bias with those targets that were expected to contain a /ʃ/ phoneme, compared to those that would usually contain a /s/ phoneme. This effect was robust in each study and always in the same direction. In Experiments 2-5, we thought this pattern of results could have been a consequence of only using a single pair of target words, with the effect representing individual differences between the words being tested. In this case we would not expect the effect to generalise across further target words. However, even in Experiment 6, when multiple target words were employed, this effect remained. This was especially perplexing as during the pre-test participants rated the /s/ and /ʃ/ targets with a difference of only 0.01 in the proportion of 'words' values.

A possible explanation, as argued for in Experiment 6, was that the consistent differences observed between /s/ and /ʃ/ target words was based on the experience of these phonemes that the listeners gained in the preceding sentence place holders. The /s/ phoneme occurs more frequently than /ʃ/ in English and that pattern was repeated in Experiment 6. In the sentence place holders there were 14 instances of /s/ phonemes heard, compared to only 4 instances of /ʃ/. The unbalanced experience of the speaker producing these phonemes in other contexts could have influenced the listeners' sensitivity to what was an acceptable /s/ versus what was an acceptable /ʃ/. This explanation would account for why there were limited differences when the targets were presented as single words in the pre-test for Experiment 6, compared to the reliable differences seen between targets in the main study when they were heard following a sentence place holder. In retrospect, this pattern of an increased number of /s/ phonemes being heard in sentence place holders in comparison to /ʃ/ phonemes held for all previous speech perception experiments in the current thesis.

This may implicate the inclusion of sentence place holder in our paradigms as having driven the target effect observed. Although this unbalanced pattern of /s/ and /ʃ/ occurrence mirrors what occurs in everyday speech, if the target effect is indeed being driven by a listeners' experience of the each phoneme in the preceding utterance, then this highlights the sensitivity of the speech perception system to the surrounding linguistic content based on the demands of the task. In future studies, this phoneme occurrence in the sentence place holder could be balanced to exert experimental control. This would test whether this target effect is being driven by this experience factor.

As argued previously, additional possible explanations for the effect of target centre on factors that we cannot investigate from the current sets of results such as the difference in the acoustic characteristics of the stimuli from Pitt and Szostak's study, differences between British and American English or sensitivity to variation that may vary between groups of participants from these respective countries. If our continuum was more natural sounding than Pitt and Szostak's then this could have influenced participants' sensitivity to the phoneme variation heard. It is worth noting that we do not know how Pitt and Szostak's pattern of results breaks down for targets expected to contain an /s/ or /ʃ/ phoneme but regardless of any differences between targets, their results generalised across the full continuum.

In Experiment 6 there was an interesting effect observed showing that focused attention and disfluency effects appeared to be having unbalanced effects on targets words. Focused attention caused a much larger effect on the /s/ target words, whereas, the opposite was seen for disfluency, with it showing larger effects on the /ʃ/ target words. We argued above that a possible explanation for these effects relates to the target effect seen consistently throughout the speech perception studies; if participants were being more accommodating of fricative variance when they were expecting to hear an /ʃ/ phoneme, even if this effect was driven by experience of each phoneme in the preceding sentence place holder, it follows that we could categorise the /ʃ/ targets as consistently being heard as more 'word' like. This would create

unbalanced lexical biases for the same fricative variation between targets, with an increased amount for /ʃ/ target compared to the /s/ targets. Therefore, we suggested that the incidence of disfluency was having a facilitative effect on lexical bias, which then increased the lexical bias to a higher level for the /ʃ/ targets, compared to the /s/ targets for the same continuum point, leading to an increase in 'word' responses from participants.

However, focused attention was acting in the predicted direction for the /s/ targets, resulting in a decrease compared to the unfocused instructions with only a very small reduction seen between the /ʃ/ targets. We proposed here that for the /s/ phonemes that as participants already had a reduced lexical bias for these targets, the added sensitivity to the bottom-up processing of the phoneme variation from the focused attention led to decrease in proportion of word values seen. A smaller effect is seen for /ʃ/ targets because these targets started with increased lexical bias, the effect of increased sensitivity to the phoneme variation counteracts the lexical bias but to a level where participants still have an increased tendency for 'word' responses.

An explanation for the variance in the impact of lexical bias between the /s/ and /ʃ/ targets would be a cognitive load effect. If after encountering disfluency there was an increase in attention, this could lead to an increase in cognitive load experienced by the participant. Under cognitive load participants have been shown to rely more on lexical influences (Mattys & Wiget, 2011). It follows that the strength of lexical bias could drive a variable impact. This would provide support for a variable attentional mechanism as proposed by Mirman, McClelland, Holt and Magnuson (2008) who suggest that it be added to interactive models of speech perception to account for the variable perceptual processes that occur due to either implicit task demands or explicit focused attention. These results support the idea of variable processing based on task demands, which has implications for the mechanisms for disfluency processing and the related accounts that are discussed below.

8.2.4 Implications for accounts of disfluency processing

In the initial investigation of predictional versus attentional accounts of disfluency processing using the eye-tracking paradigm, the results did not support either of these accounts. We suggested that the pattern of results observed could be supported by the idea of a combined account that is reliant on the integration of variable bottom-up and top-down processing dictated by task demands.

Similarly in the speech perception experiments, our exploration of the role of attention in disfluency processing provided us with unexpected results that showed that encountering disfluency made participants more accommodating of the fricative variation heard (Experiment 5 & 6), whereas, focused attention made participants more sensitive to the phoneme variation heard (Experiment 4 & 6). In our final study, the effects of both disfluency and focused attention could be seen concurrently. Taken together, these results suggested that disfluency was still having an impact that was consistent with increased attention but that it was not being directed to bottom-up processing, instead it was enhancing the top-down lexical bias and that these two levels of processing could co-occur. Again these findings could be supported by a combined account.

Although not a central focus of the current study, it is pertinent to briefly discuss the results of the current thesis in relation to the Temporal Delay Hypothesis detailed in the literature review (Chapter 2) and how this account could provide knowledge about the parameters that can affect disfluency processing. The crucial question here is whether the results of the current research are informative about disfluency per se or may instead be driven by the additional time available for processing in disfluent conditions.

The results seen for the eye-tracking paradigm employed in Experiment 1 are hard to distinguish from a temporal delay explanation as there was not a silent pause condition that employed a delay of the same duration without the incidence of the filled pause. The pattern showing an increased proportion of looks towards the

'predicted' item in the disfluent could then explained by the allowance of extra time processing for linguistic top-down expectancy for the upcoming referent in the scene.

However, the results seen in the speech perception paradigm in Experiment 5 cannot be reconciled with a simple temporal delay explanation as the variable impact of disfluency seen at Continuum point 2 and 3 would not be predicted. If the effect was being driven by only a temporal delay, then it would follow that the disfluency effect would be consistent across all continuum points, as both variants of filled pause used were of consistent length. It is hard to account for a filled pause having a variable impact at different continuum points unless the disfluency processing was also variable, which the temporal delay hypothesis alone cannot explain. A partial temporal delay explanation cannot be completely ruled out as again there was no condition that compared the filled pause disfluency condition with a condition that included a silent pause of the same duration.

Taken together the results of the current thesis suggest that a temporal delay may have had an impact upon the processing of concurrent linguistic information during comprehension but the variable disfluency effect observed during the speech perception paradigm used in Experiment 5 does not support an explanation solely based on the temporal delay offered by the incidence of a filled pause in the disfluent condition.

The Combined account proposed here suggests that the heightened attention seen following a disfluency (e.g., Collard, Corley, MacGregor, & Donaldson, 2008) can work in an additive manner with other processing such as predictional effects (e.g., Arnold et al., 2007; Heller et al., 2014). This standpoint suggests the increased attentional resources can allow listeners to attend to either top-down or bottom-up processing based on the situational need created by the context of the utterance or the demands of the task to maximise the chance for successful comprehension. In

short, the processing undertaken following the heightening of attention after encountering disfluency is modulated by task demands.

The prioritisation of either top-down versus bottom-up processing following disfluency is dependent on the parameters that would lead to maximal efficiency for the listener in the successful resolution of the task and, furthermore, transfer of the necessary linguistic information to do this with optimal processing.

If the demands of a task required participants to listen for sub-lexical phonemic detail in order to complete the task, such as if they had to press a button upon hearing a certain phoneme, then post-disfluency, bottom-up processing, which would increase attending to the fine-acoustic detail of the incoming linguistic information, would be beneficial in completing the task and likely cause listeners to attribute the increased attentional resource post-disfluency in a bottom-up manner.

Experiment 6 provided evidence that when the listeners were given the focused instruction condition this impacted their use of the attentional mechanism towards bottom-up processing, as the knowledge that they had about the phoneme variation added an extra element to the task demands, emphasising the phonemic variation more than in the unfocused instruction condition. However, after encountering disfluency in this paradigm, there was again an attentional effect but in the unfocused instruction condition this did not drive participants to increase their sensitivity to the phoneme variation, instead it acted to enhance their lexical bias as there would have been no benefit for efficiently resolving the lexical decision task with increased attending to this phoneme variation. This result does not fit with an account of disfluency processing where increased attention always drives increases in bottom-up processing for listeners. Instead, the attentional mechanism can be variably employed based on optimising comprehension for the task being undertaken.

This variable attending of incoming speech signal controlled by the demands of the task has support in the speech perception literature, where there is clear evidence for a number of different task factors that can impact perceptual processing, for

example, time course (Miller et al., 1984; Miller & Dexter, 1988) or lexical bias (Ganong, 1980; Pitt & Szostak, 2012; Pitt, 2009; Samuel, 1987), as discussed in detail in the literature above. Cognitive load is a possible driver of attentional modulation (and its impact on facilitating bottom-up or top-down processing that follows a disfluency) that is highly relevant to the combined account. However, when a task demands that focused attention is directed to the speech stream, as during a lexical decision task on ambiguous stimuli, listeners show an increased sensitivity to the bottom-up processing of fine acoustic detail (Pitt & Szostak, 2012). The variable role of attention in speech perception has been proposed (e.g., Mirman et al., 2008) but not for disfluency processing. The findings reported in this thesis lead us to propose a Combined account of disfluency processing. This account provides a framework for further exploration of and specification of the impact of cognitive load, attention and prediction on disfluency processing.

8.3 Conclusions and Future Research

Both paradigms used to explore disfluency processing in the current thesis have suggested the need for a variable mechanism for disfluency processing, one that is sensitive to both attention and task demands, maybe as a consequence of cognitive load. This variable processing adds flexibility to the comprehension of disfluency and makes use of attentional resources in a way that maximise the chance of successful message transfer.

The current thesis has highlighted that there needs to be further research to understand the complex perceptual processing that occurs following a disfluency with focus given to the role of task, attention and prediction in modulating the comprehension processes following disfluency. In future studies the next logical step would be to explicitly test the impact with and without an cognitive load task, such as in Mattys and Wiget (2011) on the lexical decision task within the speech perception paradigm used here with and without the influence of disfluency.

The combined account would predict that in a condition where a listener is subject to increased cognitive load, the incidence of a disfluency would drive increased top-down influence on comprehensive processing, as cognitive load has been shown to increase lexical influences, the post-disfluency attentional peak would enhance this in comparison to a listener who is subject to a fluent presentation whilst subject to the same cognitive load. This study would allow the effect of disfluency, cognitive load and their interaction on listeners' sensitivity to phoneme variation to be compared. This would enhance the understanding of how task demands can influence the variable impact of the observed attentional effects during disfluency processing and whether cognitive load is driving this modulation between top-down and bottom-up processing.

APPENDIX A

Instructions used in Speech perception Studies:

EXPERIMENTS 2 & 3:

Focused Attention Condition:

"Welcome and thank you for taking part.

Please listen to the speech,

There could be some mistakes. These will always be in the final word.

Changes could be small and are sound based and will be at the START of the final word. So listen carefully.

At the end of sentence you will have to select whether you thought the final word was a word or not. You should do this as quickly and accurately as possible.

You will do this by pressing a button to select the option you want:

For a WORD press the 'LEFT CTRL' key

For a NON-WORD press the 'RIGHT CTRL' key

(New Page)

There will be comprehension questions after a number of items. For these questions you will have to select one of two answers using the same buttons as above:

For the answer on the LEFT of the screen press the 'LEFT CTRL' key

For the answer on the RIGHT of the screen press the 'RIGHT CTRL' key

Please press the SPACEBAR to begin the experiment. "

Unfocused Condition:

"Welcome and thank you for taking part.

Please listen to the speech,

There could be some mistakes.

At the end of sentence you will have to select whether you thought the final word was a word or not. You should do this as quickly and accurately as possible.

You will do this by pressing a button to select the option you want:

For a WORD press the 'LEFT CTRL' key

For a NON-WORD press the 'RIGHT CTRL' key

(New Page)

There will be comprehension questions after a number of items. For these questions you will have to select one of two answers using the same buttons as above:

For the answer on the LEFT of the screen press the 'LEFT CTRL' key

For the answer on the RIGHT of the screen press the 'RIGHT CTRL' key

Please press the SPACEBAR to begin the experiment. "

EXPERIMENTS 4, 5 & 6:

Focused Instructions:

"Welcome and thank you for taking part.

Please listen to the following speech. It is of a native English speaker giving instructions about a word list.

There could be some mistakes; these will always be in the final word.

Changes could be small and are sound based. They will be located in the middle of the final word. The *S* or *SH* letter sounds can be ambiguous in word medial positions.

At the end of sentence you will have to select whether you thought the final word was a word or not.
So please listen carefully so you can make the correct response.

You will do this by pressing a button to select the option you want:

For a WORD press the 'RIGHT CTRL' key.

For a NON-WORD press the 'LEFT CTRL' key.

You should do this as *quickly* and *accurately* as possible.

(New Page)

There will be comprehension questions after a number of items. For these questions you will have to select one of two answers using the same buttons as above:

For the answer on the LEFT of the screen press the 'LEFT CTRL' key

For the answer on the RIGHT of the screen press the 'RIGHT CTRL' key

Please press the SPACEBAR to begin the experiment. "

Unfocused Instructions:

"Welcome and thank you for taking part.

Please listen to the following speech. It is of a native English speaker giving instructions about a word list.

There could be some mistakes; these will always be in the final word.

Changes could be small and are sound based.

At the end of sentence you will have to select whether you thought the final word was a word or not.
So please listen carefully so you can make the correct response.

You will do this by pressing a button to select the option you want:

For a WORD press the 'RIGHT CTRL' key.

For a NON-WORD press the 'LEFT CTRL' key.

You should do this as *quickly* and *accurately* as possible.

(New Page)

There will be comprehension questions after a number of items. For these questions you will have to select one of two answers using the same buttons as above:

For the answer on the LEFT of the screen press the 'LEFT CTRL' key

For the answer on the RIGHT of the screen press the 'RIGHT CTRL' key

Please press the SPACEBAR to begin the experiment. "

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*, 419–439.

Altmann, G. T. ., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. doi:10.1016/j.jml.2006.12.004

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 914–30. doi:10.1037/0278-7393.33.5.914

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578–582.

Bailey, K., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*(2), 183–200. doi:10.1016/S0749-596X(03)00027-5

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1235–1243.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In S. Santi, I. Gu˜atella, C. Cave, & G. Konopcyznski (Eds.), Oralit´e et gestualit´e: Interactions et comportements multimodaux dans la communication (pp. 597–600). L'Harmattan.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455. doi:10.1080/01690960903047122

Beattie, G., & Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*. Retrieved from http://las.sagepub.com/content/22/3/201.short

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, *113*(2), 1001–1024.

Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.14, retrieved 24 July 2015 from http://www.praat.org/

Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The Temporal Unfolding of Local Acoustic Information and Sentence Context, *29*(2), 155–168.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998a). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*(5), 2670–2676. doi:10.1121/1.422787

Borsky, S., Tuller, B., & Shapiro, L. P. (1998b). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*(5 Pt 1), 2670–2676.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender *. *Language and Speech*, *44*(Pt 2), 123–147.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native "um"s elicit prediction of low-frequency referents, but non-native "um"s do not. *Journal of Memory and Language*, *75*, 104–116. doi:10.1016/j.jml.2014.05.004

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Linguistics and Language Compass*, *8*(2), 37–50.

Brennan, M., & Schober, S. (2001). How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, *44*(2), 274–296. doi:10.1006/jmla.2000.2753

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398.

Christenfeld, N. (1995). Does it hurt to say um? *Journal of Nonverbal Behavior*, *19*(3), 171–186.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–42. doi:10.1006/cogp.1998.0693

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, *64*(1), 44–56.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: evidence from ERPs. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(3), 696–702. doi:10.1037/0278-7393.34.3.696

Connine, C. M. (1987). Constraints on Interactive Processes in Auditory Word The Role of Sentence Context Recognition. *Journal of Memory and Language, 26*, 527–538.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: the role of variant frequency. *Perception & Psychophysics*, *70*(3), 403–411.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. doi:10.1016/0010-0285(74)90005-X

Corley, M. (2010). Making predictions from speech with repairs: Evidence from eye movements. *Language and Cognitive Processes, 25* (5), 706-727.

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: the temporal delay hypothesis. *PloS One*, *6*(5), e19792. doi:10.1371/journal.pone.0019792

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–68. doi:10.1016/j.cognition.2006.10.010

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Linguistics and Language Compass*, *2* (4), 589-602.

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(5), 1057–1074.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141-177.

DeLong, K. a, Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–21. doi:10.1038/nn1504

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–79. doi:10.1146/annurev.psych.55.090902.142028

Dikker, S., & Pylkkanen, L. (2011). Before the N400: effects of lexical-semantic violations in visual cortex. *Brain and Language*, *118*(1-2), 23–8. doi:10.1016/j.bandl.2011.02.006

Dikker, S., Rabagliati, H., Farmer, T. a, & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, *21*(5), 629–34. doi:10.1177/0956797610367751

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*(3), 293–321. doi:10.1016/j.cognition.2008.09.008

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436.

Eklund, R. (2001). Prolongations : A dark horse in the disfluency stable, 5–8.

Eklund, R., & Shriberg, E. (1998). Crosslinguistic Disfluency Modeling: A Comparative Analysis of Swedish and American English Human-Human and Human-Machine Dialogs. In *Fifth International Conference on Spoken Language Processing* (pp. 1–4).

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. doi:10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. doi:10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. doi:10.1016/j.brainres.2006.06.101

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*(2), 233–253.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*(8), 1151–1177. doi:10.1080/01690960701461293

Finlayson, I. R. (2014). *Testing the roles of disfluency and rate of speech in the coordination of conversation.* Unpublished PhD Dissertation, Queen Margaret University.

Finlayson, I. R. & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review, 19,* 921-928. doi:10.3758/s13423-012-0279-x

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738. Retrieved from http://www.sciencedirect.com/science/article/pii/S0749596X85710327

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, *29*(2), 320–326.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151–167.

Ganong, W. F. (1980). Phonetic Categorization in Auditory Word Perception, *6*(1), 110–125.

Gaskell, M. G., & Marslen-wilson, W. D. (1998). Mechanisms of Phonological Inference in Speech Perception, *24*(2), 380–396.

Goldinger, S. D. (1999). Only the shadower knows: comment on Hamburger and Slowiaczek (1996). *Psychonomic Bulletin & Review*, *6*(2), 347–351; discussion 352–355.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*(3), 169–77. doi:10.1027/1618-3169/a000021

Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (2014). Inferring Difficulty: Flexibility in the Real-time Processing of Disfluency. *Language and Speech*. doi:10.1177/0023830914528107

Hieke, A. E. (1981). A Content-Processing View of Hesitation Phenomena. *Language & Speech*, *24*(2), 147–160. Retrieved from http://search.ebscohost.com.simsrad.net.ocs.mq.edu.au/login.aspx?direct=true&db=aph&AN=14091125&site=ehost-live

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*(4), 305–12. doi:10.1111/j.0956-7976.2005.01532.x

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *The Journal of the Acoustical Society of America*, *119*(6), 4016. doi:10.1121/1.2195119

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *The Journal of the Acoustical Society of America*, *108*(2), 710. doi:10.1121/1.429604

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical …*, *94*(2), 701–714. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/94/2/10.1121/1.406887

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Kamide, Y., Altmann, G. T. ., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. doi:10.1016/S0749-596X(03)00023-8

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. doi:10.1016/j.cognition.2004.03.002

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81. doi:10.1016/j.cognition.2007.07.013

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, *121*(3), 459–65. doi:10.1016/j.cognition.2011.08.015

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: how listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–8. doi:10.1111/j.1467-9280.2008.02090.x

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*.

Kutas, M., & Hillyard, S. A. (n.d.). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kwon, N., & Sturt, P. (2014). The use of control information in dependency formation: An eye-tracking study. *Journal of Memory and Language*, *73*, 59–80. doi:10.1016/j.jml.2014.02.005

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/29/1/10.1121/1.1908694

Laing, E. J. C., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a Tune: Talker Normalization via General Auditory Processes. *Frontiers in Psychology*, *3*(June), 203. doi:10.3389/fpsyg.2012.00203

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*(1), 74–88. doi:10.1016/j.bandl.2006.02.003

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. doi:10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–77. doi:10.1016/j.cognition.2007.05.006

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*, 1–55.

Lickley, R. J., & Bard, E. G. (1996). On not recognizing disfluencies in dialogue. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, *3*.

Lindblom, B. E., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America*, *42*(4), 830–843.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, *48*(14), 3982–3992.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*, 19–44.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science*.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412.

Marslen-Wilson, W. (1989*). Access and integration: Projecting sound onto meaning.* Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*.

Martin, J. G., & Strange, W. (1968). Determinants of Hesitations in Spontaneous Speech. *Journal of Experimental Psychology*, *76*(3, Pt.1), 474–479. doi:10.1037/h0025598

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145–160. doi:10.1016/j.jml.2011.04.004

Mcclelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech approach , information pro-, *86*.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–9. doi:10.1016/j.tics.2006.06.007

McQueen, J. (2003). The ghost of Christmas future: didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science*, *27*(5), 795–799. doi:10.1016/S0364-0213(03)00069-7

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(2), 433–443.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language*, *61*(1), 1–18. doi:10.1016/j.jml.2009.03.002

Miller, J. L., & Dexter, E. R. (1988). Effects of Speaking Rate and Lexical Status on Phonetic Perception, *14*(3), 369–378.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics*, *36*(4), 329–337.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language*, *52*(3), 416–435. doi:10.1016/j.jml.2005.01.006

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of Attention on the Strength of Lexical Influences on Speech Perception: Behavioral Experiments and Computational Mechanisms. *Cognitive Science*, *32*(2), 398–417. doi:10.1080/03640210701864063

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *The Behavioral and Brain Sciences*, *23*(3), 299–325; discussion 325–370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*.

O'Connell, D. C., & Kowal, S. (2005). Uh and um revisited: are they interjections for signaling delay? *Journal of Psycholinguistic Research*, *34*(6), 555–76. doi:10.1007/s10936-005-9164-3

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. doi:10.1121/1.1906875

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–10. doi:10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–47. doi:10.1017/S0140525X12001495

Pinnow, E., & Connine, C. M. (2013). Phonological Variant Recognition: Representations and Rules. *Language and Speech*, *57*(1), 42–67. doi:10.1177/0023830913479105

Pitt, M. a. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(3), 896–910. doi:10.1037/a0013160

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, *61*(1), 19–36.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*(1), 89–95.

Pitt, M. A., & Szostak, C. M. (2012). A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes*, *27*(7-8), 1225–1239. doi:10.1080/01690965.2011.619370

Plauché, M., & Shriberg, E. (1999). Data-driven subclassification of disfluent repetitions based on prosodic features. *Proc. International Congress of Phonetic …*, 1–4. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.367&rep=rep1&type=pdf

Ranbom, L., & Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. doi:10.1016/j.jml.2007.04.001

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, Available: http://www.R-project.org/. ISBN 3-900051-07-0.

Rohde, H., & Ettlinger, M. (2012). Integration of pragmatic and phonetic cues in spoken word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(4), 967–83. doi:10.1037/a0026786

Sajin, S. M., & Connine, C. M. (2014). Semantic richness: The role of semantic features in processing spoken words. *Journal of Memory and Language*, *70*, 13–35. doi:10.1016/j.jml.2013.09.006

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, *105*(2), 466–476.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474–494. doi:10.1037//0096-3445.110.4.474

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, *26*(1), 36–56. doi:10.1016/0749-596X(87)90061-1

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science : A Journal of the American Psychological Society / APS*, *12*(4), 348–351.

Samuel, A. G., & Ressler, W. H. (1986). Attention within auditory word perception: insights from the phonemic restoration illusion. *Journal of Experimental Psychology. Human Perception and Performance*, *12*(1), 70–79.

Schnadt, M. J. (2009). *Lexical influences on disfluency production.* Unpublished PhD Dissertation, University of Edinburgh.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499–520.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*(2), 232–252. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0749596X85900269

Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*(3), 330–346.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.

Tuinman, a., Mitterer, H., & Cutler, a. (2013). Use of Syntax in Perceptual Compensation for Phonological Reduction. *Language and Speech*, *57*(1), 68–85. doi:10.1177/0023830913479106

Van Alphen, P., & McQueen, J. M. (2001). The time-limited influence of sentential context on function word identification. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(5), 1057–1071.

Van Berkum, J. J. a, Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 443–67. doi:10.1037/0278-7393.31.3.443

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2).

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. doi:10.1162/0898929041920487

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, *28*(3), 272–302. doi:10.1080/01690965.2011.622905